DOCUMENT RESUME

ED 062 346 TE 002 898

AUTHOR Follettie, Joseph F.; Wesemann, Ann F.
TITLE Effects of Grammatical Factors and Amount of Material on Memorizing Paragraphs, Sentences, and Word Lists.
INSTITUTION George Washington Univ., Alexandria, Va. Human Resources Research Office.
SPONS AGENCY Department of the Army, Washington, D.C.
REPORT NO TR-67-9
PUB DATE Jun 67
NOTE 72p.; Basic Research Study 7

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Grammar; Instructional Materials; *Memory; *Paragraphs; Recall (Psychological); Semantics; *Sentences; Standards; Stimuli; Time Factors (Learning); *Word Lists

ABSTRACT

Research conducted to determine the effect of certain measurable characteristics of written material upon speed of memorization is presented. The characteristics studied fall into four classes: (1) Load measures reflecting informational density, (2) Length measures based upon number of syllables, words, kernel sentences, clauses, or sentences in a passage, (3) Packaging measures based on alternative, gramatically equivalent ways for dealing with the same semantic material, and (4) Word frequency measures. Six studies comprising 14 experiments were conducted. Experimental materials consisted of paragraphs, lists of sentences, single sentences, and lists of words. In all but one of the six studies, the primary measure of learning was cumulative presentation time to a criterion of one perfect rote recall of the experimental material, where presentation time during each exposure of the material was under the subject's control. Results include: (1) Paragraphs with a very high ratio between content words and total words in the paragraph required more time to learn than those with a lower content load; and (2) The time required to learn a set of sentences increased as the mean number of syllables per content word increased. Conclusions include: (1) Of the four types of stimulus characteristics studied, length has the most potent effects upon rote memorization; and (2) A number of relationships between independent and dependent variables were found to yield good fits to linear, power, or exponential equations. (Author/CK)

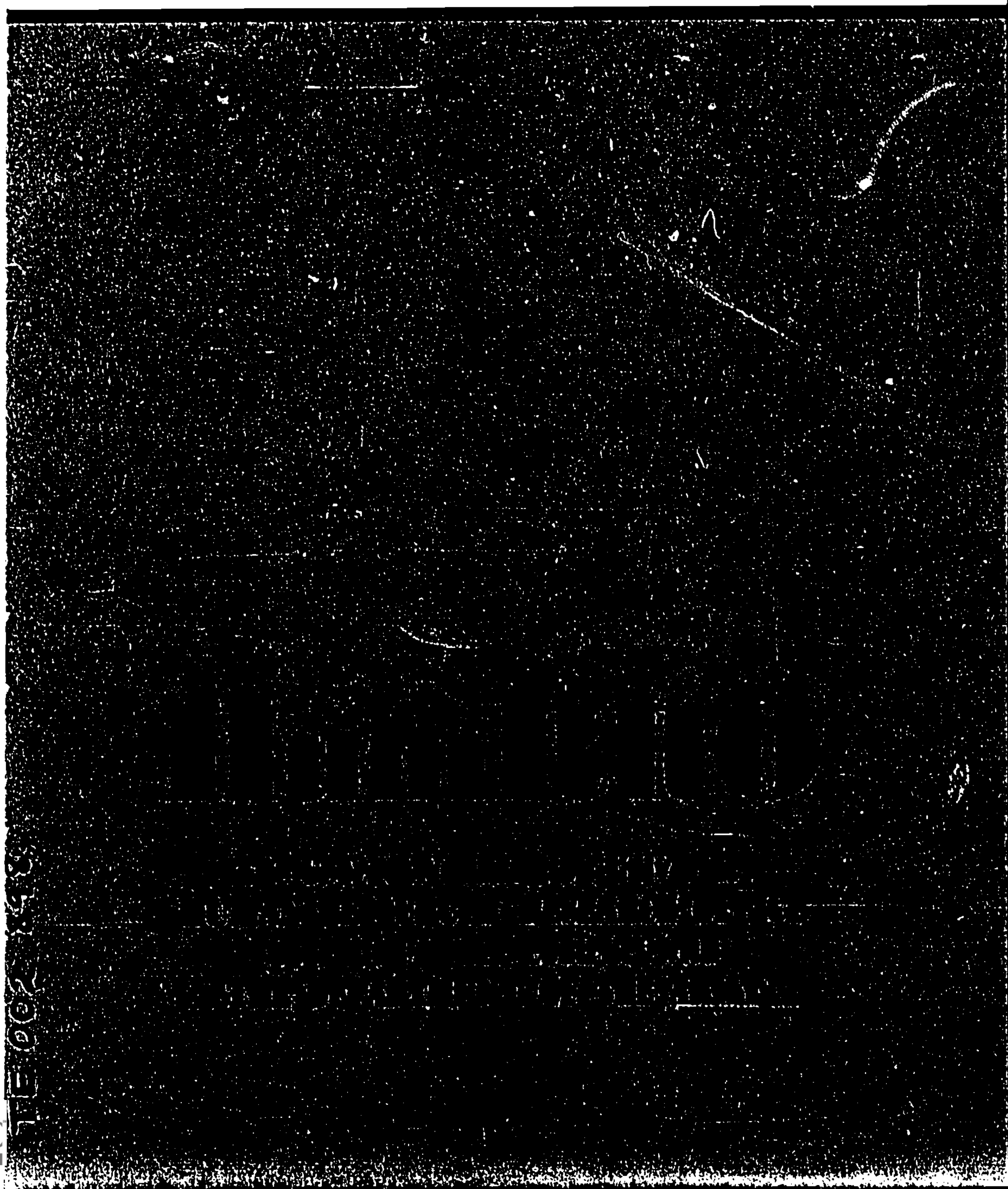
Technical Report 67-9

Effects of Grammatical Factors and Amount of Material on Memorizing Paragraphs, Sentences, and Word Lists

by

Joseph F. Follettie and Ann F. Wesemann

HumRRO Division No. 5 (Air Defense)



June 1967

Prepared for:

Office, Chief of Research and Development Department of the Army

Contract DA 44-188-ARO-2

Distribution of this document is unlimited.

ED 062346

Effects of Grammatical Factors and Amount of Material on Memorizing Paragraphs, Sentences, and Word Lists

by

Joseph F. Follettie and Ann F. Wesemann

Distribution of this document is unlimited.

June 1967

Prepared for:
Office, Chief of Research and Development
Department of the Army
Contract DA 44-188-ARO-2 (DA Proj 2J014501B74B 02)

HumRRO Division No. 5 (Air Defense)
Fort Bliss, Texas
The George Washington University
HUMAN RESOURCES RESEARCH OFFICE
operating under contract with
THE DEPARTMENT OF THE ARMY

Technical Report 67-9
Basic Research
Study 7

The Human Resources Research Office is a nongovernmental agency of The George Washington University, operating under contract with the Department of the Army (DA 44-188-ARO-2). HumRRO's mission is to conduct research in the fields of training, motivation, and leadership.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Published
June 1967
by
The George Washington University
HUMAN RESOURCES RESEARCH OFFICE
300 North Washington Street
Alexandria, Virginia 22314

Distributed under the authority of the
Chief of Research and Development
Department of the Army
Washington, D.C. 20310

FOREWORD

The studies described in this report were undertaken as part of the research conducted by the Human Resources Research Office under Basic Research Study 7, Precision of Statement and Perception of Meaning of Written Language in Military Training. These studies relate to the problem of quantifying the difficulty of written material on technical subjects and determining empirically rules for writing that minimize the difficulty of such material.

Basic Research Study 7 was conducted by HumRRO Division No. 5 (Air Defense) at Fort Bliss, Texas, while Dr. Robert D. Baldwin was Director of Research. Military support for the study was provided by the U.S. Army Air Defense Human Research Unit. The Military Chief of the Unit was LTC Leo M. Blanchett, Jr. SP 4 Fred G. Hampton provided assistance in the data collection and analysis phases of the research.

HumRRO research is conducted under Army Contract DA 44-188-ARO-2, with the Basic Research studies under Army Project No. 2J014501B74B 02, Basic Research in Psychology and Social Science.

Meredith P. Crawford
Director
Human Resources Research Office

GLOSSARY

Content Word Ratio (CWR)	The ratio between total content words and total words in the passage.
Content words	Words commonly classified as nouns, verbs, adjectives, and adverbs, together with substitute classes of words (such as pronoun for noun).
Free recall	A response in which elements may occur in any order.
Function words	Words not classified as content words.
Kernel sentence	A grammatical unit that can be applied to text to analyze content in terms of simplest possible declarative statements that express complete thoughts; that is, in each instance, anything less becomes a non-sentence. A normal sentence is comprised of one or more kernel sentences.
Length	Number of sentences, clauses, kernel sentences, words and syllables have all been used as measures of length.
Linear relationship	Form of relationship between two variables that can be represented by a straight line (or the equation for a straight line: $Y = aX + b$), where X and Y are variables and a and b are constants.
Load	An index of information; the information density of a unit of content of fixed length.
Monotonic relationship	Form of relationship in which direction of change in value in one variable is associated with uniformity of direction of change for the other. That is, the value for one variable (either increasing or decreasing) is associated with a uniform *direction* of change in the other (either increasing or decreasing) over the range of both variables. A linear relationship is the simplest form of a monotonic relationship. A relation in which there are cyclic changes in a variable with increasing values of another variable illustrates functions that are *not* monotonic.
Packaging	Alternative construction for dealing with the same semantic material.
Rote recall	A response in which elements must occur in the order in which they occur in the stimulus.
Tokens	*All* words in a passage (see Type-Token Ratio).
Trial	Fixed-length exposure of a unit of content.
"Trial"	Variable-length (subject-controlled) exposure of a unit of content.
Types	*Different* words in a passage (see Type-Token Ratio).
Type-Token Ratio (TTR)	The ratio of the number of different words (types) to this total number of words (tokens) in a passage.
Word frequency	A word's frequency in print, as reported by one of the counts in Reference **18**.

SUMMARY AND CONCLUSIONS

Problem

The research conducted under Basic Research Study 7 of the Human Resources Research Office bears upon two related problems. The first problem is that of determining how to evaluate the difficulty of technical written material; for any example of written material dealing with a technical subject, it would be desirable to be able, eventually, to assign a number to the material which would indicate where it falls on a scale of difficulty. The second problem is that of determining what can be done to make given material less difficult. If a writer has two or more options as to how he can express a proposition, idea, relationship, and so forth, which option should he elect in order to minimize the difficulty a reader will experience in trying to understand the writing?

The present studies were conducted to determine the effect of certain measurable characteristics of written material upon speed of memorization.[1] While memorizing and understanding are not the same process, memorization measures—which are quicker, more reliable, and easier to use than measures of understanding—were used in these experiments on the assumption that such measures would be predictive of understanding.

The characteristics studied in these experiments fall into four general classes:

(1) *Load measures* reflecting informational density.
(2) *Length measures* based upon number of syllables, words, kernel sentences, clauses, or sentences in a passage.
(3) *Packaging measures* based on alternative, grammatically equivalent ways for dealing with the same semantic material.
(4) *Word frequency* (familiarity) measures.

Research Procedure

Six studies comprising 14 experiments were conducted. Experimental materials consisted of paragraphs, lists of sentences, single sentences, and lists of words.

In all but one of the six studies, the primary measure of learning was cumulative presentation time to a criterion of one perfect rote recall of the experimental material, where presentation time during each exposure of the material was under the subject's control. In the sixth study, three experiments employed the same measure, but to a criterion of one perfect free recall of the material. In two other experiments, exposure time on each trial was fixed and thus not under the subject's control; a trials-to-criterion measure of learning speed was used in these experiments.

Results

(1) Paragraphs with a very high ratio between content words (nouns, verbs, adjectives, and adverbs) and total words in the paragraph required more time to learn than those with a lower content word load. However, content word load variations within or below the normal range for written English had negligible effects on learning time.

(2) The time required to learn a set of sentences increased as the mean number of syllables per content word increased.

[1]As a convenience to the reader, a glossary of some technical, linguistic, and mathematical terms, defined as they are used in this report, is presented on p. iv.

6

(3) In two studies using small samples of subjects, reasonably smooth curves were found to represent the effects of list length, in sentences, upon recall. Learning time measures were differently related to list length in the two studies, apparently due to the fact that the amount of content word repetition varied with list length in one study and not in the other.

(4) A pilot study yielded results which suggest that single-sentence recall is an orderly joint function of number of kernel sentences in the clause and number of clauses in the sentence.

(5) In material featuring a string of serial noun phrase modifiers, placing the series before the noun phrase head, rather than after it, was found to facilitate single-sentence recall. When modifiers were coordinate, however, learning was about the same regardless of whether the modifiers were placed before or after the head. Where pre-head placement led to quicker learning, the result can be accounted for entirely on the basis of sentence length (in syllables), since the post-head position of the modifiers required use of more words in the sentence.

(6) Studies on the joint effects of word length and word frequency on the recall of word lists indicate that presentation time to criterion is a monotonic, increasing function of word length. Simple frequency effects and joint effects of length and frequency varied from one study to another with changes in procedure and in the particular lists used. In all such studies, mean response rate increased with increasing word frequency.

Conclusions

(1) Of the four types of stimulus characteristics studied—load, length, packaging, and frequency—length has by far the clearest and most potent effects upon rote memorization.

(2) A number of relationships between independent and dependent variables were found to yield good fits to linear, power, or exponential equations. Results of this type may have application in the eventual development of mathematical equations that will represent and measure the level of difficulty of technical written material.

(3) The present findings have implications for the possible course of future research exploring the nature of technical language difficulty. Some of these implications are discussed in Chapter 8.

CONTENTS

Effects of Grammatical Factors and Amount of Material on Memorizing Paragraphs, Sentences, and Word Lists

Chapter 1

INTRODUCTION

This report describes a series of experimental studies, conducted in Basic Research Study 7 of the Human Resources Research Office, exploring language factors that may influence the difficulty of technical written material. The general objectives of the research in this area are:

(1) To predict the effort that a reader will need to expend in order to understand given written material. (How difficult is the material?)

(2) To indicate how such material can be modified to minimize its superficial complexity so that the reader will need to expend less effort to understand it. (What can be done to make the material less difficult?)

BACKGROUND

The research stems from two primitive, or basic, notions. First, written material has measurable characteristics that determine its level of complexity. Second, the level of complexity of a passage determines how difficult it will be for a given reader to understand. Studies predicated on these notions have been under way in one form or another for 40 years. The earliest of these investigations can be traced to the practical requirement in public school education that textbooks be written so school children at particular grade levels can understand them.

At least three trends in educational and psychological research appear in some degree relevant to the present work. The earliest studies sought to predict the readability of written material on the basis of correlations between reading test scores and certain gross statistical and grammatical characteristics of the material. Such studies, leading to the production of readability formulas, appear with regularity in the literature.

The studies reflecting a second trend—experimental rather than correlational in approach—seek to predict the understandability of written material in consequence of deliberate manipulation of certain characteristics of the material. There are fewer studies of this sort. Most of them are recent and, as a result, some of them make use of recent advances in the theory of grammar to define characteristics of the material which will be manipulated (a tendency not yet evident in the readability research).

A third research trend relates to language development in children. Studies on this topic traditionally have addressed the proposition that as children mature, their writing will become linguistically more complex. A recent tendency in language development research has been to focus on the grammatical complexity of the unit of reading and its relation to developmental levels.

While we do not intend here to review the literature associated with these trends, perhaps it would be useful to describe their methodologies.

Correlational Research. Correlational research has tended almost exclusively to be readability research. The readability formula is a device

used by educators to evaluate the relative difficulty of textbooks. These formulas typically are regression equations based upon correlation studies. At the outset of such a study, a series of passages of given length is evaluated for several characteristics, such as average length of sentences, number of dependent clauses, number of prepositional phrases, or number of repeated words. Then, subjects read the passages and are tested for understanding. Next, test scores are correlated with the various passage characteristics. Finally, those characteristics which correlate most highly with test scores (and not too highly with each other) are used as factors in readability formulas.

The readability research has been reviewed extensively (1, 2, 3, 4). We will be content here to give an indication of the kinds of factors employed in readability formulas.

Perhaps the most comprehensive research of this sort was reported 30 years ago by Gray and Leary (5). The effects of approximately 70 characteristics of written material upon the understanding of adults of limited reading ability were determined by correlational methods. A series of alternative readability formulas was generated on the basis of the resulting matrix of intercorrelations. According to the principal formula of Gray and Leary, test score is a function of five factors:

(1) Number of personal pronouns in a 100-word passage.
(2) Number of different difficult words in the passage.
(3) Average length of sentences.
(4) Percentage of different words.
(5) Number of prepositional phrases.

The household word for readability research is Flesch. According to the Flesch reading ease formula (6), reading ease score is a function of:

(1) Average sentence length of the passage, in words.
(2) Number of syllables per 100 words.

The economy, in factors used, of the Flesch formula over the Gray and Leary formula is noteworthy. The Flesch formula reflects a tendency in readability research in recent years to yield formulas based on only a few factors, of the sort that can be reliably assessed by almost any literate person. Perhaps such simple formulas are adequate when used to classify textbooks by age-grade at the public school level. Whether they are adequate for evaluating the difficulty of technical material of the sort found in college-level introductory textbooks or in the course materials of Army Technical Training programs is a question that has received only limited study.

Stevens and Stone (7) applied one of the Flesch formulas to textbooks in psychology and found, according to the formula, that the writings of the psychologist Koffka were among the easiest readings in the field. However, most psychologists would agree that Koffka's writings are monumentally difficult. We have concluded—on the basis of informal, intuitive studies like those of Stevens and Stone—that readability formulas are not sufficiently sensitive to establish the relative difficulty of technical materials written for use in adult education.

Experimental Research. While experimental and correlational approaches to evaluating the difficulty of written material may differ methodologically in a number of ways, they need so differ in only one. The passages of an experimental study are deliberately written to reflect different levels of a few characteristics of written material—one characteristic if the research design is simple, more than one if it is complex. All other characteristics of the experimental passages are held constant. The passages of a correlational study are

not deliberately written by the investigator. If they were, no doubt they would reflect the same research design considerations to which the experimental investigator reacts. Under the latter condition, results of parallel investigations employing the two approaches would amount to no more than two different ways of saying the same thing.

The passages of a correlational study typically are selected from existing written material. The selection process guarantees that the set of passages used will reflect different levels for each of a number of characteristics of written material. However, while the experimental study allows the investigator to set his levels precisely where he wishes them to be, the correlational study to some extent imposes values for different levels of a characteristic upon the investigator. Moreover, if the experimental investigator chooses to study the effects of two or more characteristics in the same study, the passages used can be made to reflect all combinations of characteristics and levels. It is just not possible to select passages from existing written material which guarantee that all combinations of characteristics and levels will be represented. Hence, it is possible to systematically evaluate the interaction between characteristics in experimental studies, and not possible to do so in correlational studies.

There have been few experimental studies dealing with difficulty as a function of characteristics of written material. In addition to those described in this report, certain studies by Coleman (8, 9) serve as examples.

Research on Language Development. The typical study of language development of school-age children employs subjects at different ages or school grades. Most studies of this kind have begun by obtaining writing samples of some length—for example, 1000 words—from the subjects. Typically, the subject is required to write a theme on a topic of his choosing. The grammatical characteristics of the resulting material are then evaluated. The final step is to isolate those grammatical characteristics of writing—for example, clause length, proportion of all words that are adjectives—which follow from the preconception that physical (chronological) growth in the child should be accompanied by linguistic growth. Such studies suggest variables worth considering in experimental research on the difficulty of written material.

The language development literature of the sort described has been reviewed by McCarthy (10) and by Harrell (11). Many of the earlier studies—like those of readability research—have been concerned with such factors as sentence length, clause length, the ratio of subordinate clauses to all clauses, and ratios between the number of words in one grammatical class to those of another (as in the ratio of adjectives to verbs).

Hunt (12) recently published a study of language development in school-age children which takes advantage of tools of analysis that the more modern grammars invite. Hunt employs the distinction made in modern grammars between surface components and deep components. In consequence, he was able to study the effects of age (school grade) on the number of kernel sentences[1] produced or the number of sentence-combining transformations represented in a single-clause or multiple-clause construction. What is interesting about such factors is that, in the context of Hunt's study, these more sophisticated factors tended to covary with school grade more closely than did such surface factors as number of sentences in the 1000-word passage.

[1]As a convenience to the reader, a glossary of some technical, linguistic, and mathematical terms, defined as they are used in this report, is presented on p. iv.

If the maturation of school-age children is accompanied by increasing linguistic sophistication in written expression, perhaps it is also accompanied by an analogous shift in reading habits, and hence, in turn, to understanding written material. When reading, what are the units of grammar—the word, the phrase, the clause—to which one attends and do these units vary with age? An initial interest in such questions is reflected in articles by Baldwin and Baum (13) and Suci (14), who are members of the Reading Research Group at Cornell University. The preliminary work at Cornell suggests that the magnitude of the unit of grammar attended to when reading increases with age-grade. It seems worth keeping in mind that the grammatical units that prove to be basic to reading will also be basic to understanding.

CHARACTERISTICS OF THE PRESENT STUDIES

The present studies, which were conducted during the period from July 1964 to October 1965, are concerned with discovering dimensions underlying the complexity of written material and determining functional relations between these dimensions and measures of proficiency following instruction. While memorizing and understanding are not the same process, memorization measures—which are quicker, more reliable, and easier to use than measures of understanding—were used in these experiments on the assumption that such measures would be predictive of understanding.

Six studies encompassing 14 experiments are described. Each experiment utilized one of four classes of units of content—paragraphs, sentence lists, sentences, or word lists. One of three primary measures of proficiency was used—cumulative presentation time to a rote recall criterion, cumulative presentation to a free recall criterion, or fixed-interval trials to a free recall criterion. The independent variables of an experiment fell into one or more of four general classes—those of load (information density), length, packaging, and word frequency.

Units of Content

Pertinent characteristics of the four classes of units of content employed in the various experiments were as follows:

(1) Paragraphs. Each unit of content was a paragraph consisting of two sentences written in a relaxed expository style. The second sentence of a paragraph was formally related to the first, usually on the basis of a rule of the form "A noun of Sentence 1 equals a pronoun of Sentence 2." Paragraphs were used as experimental content in one study.

(2) Sentence lists. Each unit of content (that is, each list of sentences) consisted of a set of simple sentences—one-clause constructions containing a single kernel sentence. For present purposes, a "kernel sentence" will be taken as an irreducible sentence. Representative types of such sentences are "N is A" (e.g., "The shovel is dirty") and "N V N P N" (e.g., "The baker gave the cookie to the shortstop").[1] The sentences of a set were arranged in a list, one below the other.

Sentence lists were used in two studies, containing four experiments. All sentences of all lists used in an experiment were of the same kernel

[1]In all content formulas used in this report, N = noun, A = adjective, V = verb, and P = preposition.

sentence type. In one experiment, the sentences of a list were formally related through the sharing of content words. In the other three experiments, the sentences of a list were formally unrelated, although written so as to convey a sense of belonging to the same topic.

(3) Sentences. Each unit of content consisted of a single sentence. In two studies (four experiments) using sentences as units, certain characteristics of syntax (sentence grammar) or of sentence length were manipulated while other characteristics were held constant.

(4) Word lists. Where word lists were used as units of content, list length was held constant while word length and word frequency were manipulated. One study, consisting of five interrelated experiments, used word lists as the unit of content.

Measures of Proficiency

All proficiency measures involved recalling material, demonstrating it had been memorized. Recall is rote when the elements of a response—words or other units—must follow in the order in which they occur in the stimulus. It is free when the response elements may occur in any order. Free recall responding characterized experiments of the study using word lists; rote recall characterized the studies using paragraphs or sentences. The studies using sentence lists required a mixture of free and rote recall; the subject had to maintain word order within sentences, but any order of sentence recall was considered correct. For these studies, responding was classified under rote recall because that seemed the principal component of the response.

The primary measures of proficiency were as follows:

(1) Subject-controlled presentation time to rote recall. In five of the six studies, the primary measure of proficiency was cumulative time to a criterion of immediate perfect total rote recall of the unit of content, where presentation time during each exposure of the unit (or "trial"[1]) was under the subject's control and hence variable over "trials."

(2) Subject-controlled presentation time to free recall. In three experiments of the sixth study, the primary measure of proficiency was cumulative time to a criterion of immediate perfect total free recall of the unit of content, where presentation time during each exposure was under the subject's control.

(3) Fixed-interval trials to free recall. In the other two experiments of the sixth study, presentation time during an exposure was fixed. The measure of proficiency was number of trials to a criterion of immediate perfect total free recall of the unit of content.

Results obtained from other measures of proficiency are scattered through the descriptions of the studies, but the foregoing measures were considered primary when the studies were designed.

Classes of Independent Variables

The terms load, length, packaging, and frequency subsume rather well the variables whose effects were studies.

Load. The words of written material may be characterized in various ways. They may, for example, be classified by grammatical category—for example,

[1]Throughout the report, the word "trials" will be used in quotation marks for variable-length exposures, and without quotation marks to denote fixed-length exposures.

content words and function words[1]; adjectives, nouns, verbs, and so forth; finite and infinite verbs; transitive and intransitive finite verbs. Ratios between such word classes are potential indices of the difficulty of written material. Boder's (15) Adjective-Verb Quotient (AVQ) is a ratio of this sort.

Words also may be classified according to whether they are used more than once in a passage ("tokens", all words; "types", different words). It is generally held that ratios based on the distinction between word types and word tokens assess vocabulary diversification in the passage. Wendell Johnson's (16) Type-Token Ratio (TTR) is an example of this sort of ratio.

The effects on difficulty of written material of many kinds of ratios such as those illustrated by AVQ and TTR either have been or could be studied. Miller (17) catalogs a fair sampling of such ratios and touches on some of the findings regarding their effects.

We use the term "load" to refer to any ratio proposed as an index of information. Load measures are meant to reflect the informational density of a unit of content of fixed length. Ratios of the sort illustrated by the AVQ are load factors; so, under certain conditions, are ratios of the sort illustrated by TTR. To the extent that equations of information theory could be made applicable to written material, they too would provide measures of load.

The load factor investigated in the present studies was the Content Word Ratio—the ratio between total content words and total words in the passage (CWR).

Length. In the present studies, number of sentences, number of clauses, number of kernel sentences, number of words, and number of syllables have been used as measures of length.

Packaging. Packaging refers to alternative constructions for dealing with the same semantic material while holding content morphemes[2] constant, as in "Turkish action" and "action of Turkey." For example, we are dealing with a packaging variable when studying effects of whether modifiers precede or follow the noun in noun phrases.

Frequency. Whereas the factors of load and length, and perhaps even packaging, apply to all classes of units of content of every magnitude, frequency is used here only with reference to words. It is used to apply to a word's frequency in print, as this is reported in one of the counts—G Count, Magazine Count—by Thorndike and Lorge (18). One study to be reported manipulated word frequency.

INVENTORY OF STUDIES

Paragraphs: Study on Time and Load Relationships (T = f (Load))

In Study 1 (Chapter 2), cumulative presentation time to criterion was studied as a function of Content Word Ratio (CWR), the ratio of content words to total words in a two-sentence paragraph containing 30 words.

[1]See discussion in "Inventory of Studies."

[2]Linguists tend to divide words into content and function word classes. A content word contains one or more content morphemes—the stem or root component. It may or may not contain function morphemes—for example, a derivational or inflectional ending. The underlined portions of the following content words are content morphemes: *act*ive, *beaut*ify, *count*ing, *dead*ly, *end*ed, *full*, *mot*ion, *new*ish, *opera*s, *prim*ary, *raw*ness, *siz*able, *treat*ment.

Sentence Lists: Studies on Time and Length Relationships (T=f (Length))

In Study 2 (Chapter 3), cumulative presentation time to criterion was studied as a function of number of sentences to be recalled:

Experiment 2a—Sentences were formally interrelated through the sharing of content words.

Experiment 2b—Sentences were formally unrelated although written to convey a sense of belonging to the same topic.

In Study 3 (Chapter 4), cumulative presentation time to criterion was studied as a function of number of syllables in the content words of a three-sentence set:

Experiment 3a—All content words in a given set of sentences constituting the unit of content were of the same length.

Experiment 3b—The number of syllables in the content words of a unit of content varied from word to word around a mean value for syllabic length.

Sentences: Studies on Time and Length Relationships (T=f (Length))

In Study 4 (Chapter 5), cumulative presentation time to criterion was studied as a function of number of clauses in the sentence and number of kernel sentences in the clause, where clause and kernel sentenceforms were controlled.

Experiment 4a—Clauses were joined by and.

Experiment 4b—Clauses were joined by if-then.

Study 4 is a pilot study with a meager data base. Its results, however, are interesting, and since the study is prototypic of planned work, it is reported in some detail.

Sentences: Studies in Time and Packaging Relationships (T=f (Packaging))

In Study 5 (Chapter 6), time to criterion was studied as a function of the order of modifiers with respect to their noun head, with modifiers either preceding or following their head according to the syntactic conventions of English:

Experiment 5a—Modifiers were coordinate adjectives, and the head was a noun.

Experiment 5b—Modifiers were noun adjuncts or other attributives of serial modification, and the head was a noun.

Word Lists: Studies of Relationships of Time or Trials With Length and Frequency (T=f (Length, Frequency))

In Study 6 (Chapter 7), proficiency was studied as a function of number of syllables per word and word frequency, where the task was to memorize a 15-word list:

Experiments 6a, 6b, 6c—The measure of proficiency was time to criterion.

Experiments 6d, 6e—The measure was number of fixed-interval trials to criterion.

Chapter 2

PARAGRAPH RECALL AS A FUNCTION OF CONTENT WORD LOAD
(Study 1)

Abstract of Study 1

This study explored the extent to which the time needed to memorize a paragraph is affected by the ratio between content words [1] and total words in the material being memorized. The 30-word paragraphs used in the experiment contained 25, 20, 15, or 10 content words; the Content Word Ratio (CWR) was, respectively, .83, .67, .50, or .33.

Paragraphs with a high proportion of content words (CWR .83) required significantly more time to recall than those with a lower CWR. However, the differences between recall times for the paragraphs having lower CWRs (.33, .50, and .67) were not significant. Since mean recall time did not progressively increase with increasing CWR, these data are not consistent with the seemingly reasonable view that recall is directly related to the ratio of content to total words.

BACKGROUND

There are a number of reasons for expecting that increasing content word load would be accompanied by increasing difficulty in learning meaningful material. Compared with functional words, content words tend to be longer and to occur less often (Fries, 19; Miller, Newman, and Friedman, 20). Using a cloze procedure[2] to assess sources of contextual constraint in sentences, Aborn, Rubenstein, and Sterling (22) found content words to be about one-half as predictable as function words; the size of the difference in predictability was not affected by the amount of context around the omitted word. Taylor's (21) analysis of cloze data also indicates that all classes of content words except pronouns yield lower cloze scores than do function words. Since there is a high negative correlation between cloze scores and relative uncertainty (entropy), we assume that in general the information load will rise as the Content Word Ratio (CWR) increases in a prose passage.

In consequence of word frequency, word length, and information load effects obtained by other investigators, it seems tenable that passages with a high CWR should be more difficult to learn than passages with lower CWR values.

[1]While many grammars distinguish between content words and function words, the criteria for classification vary. We follow Fries (19), in whose system most words commonly classified as nouns, verbs, adjectives, and adverbs, together with substitute classes of words (e.g., pronouns for nouns) are content words and all others are function words.

[2]When several subjects guess the identity of one or more deleted words in a passage on the basis of clues furnished by context, the resulting average proportion of correct responses is called a cloze score and the general procedure a cloze procedure. The terminology traces back to Taylor (21).

METHOD

Materials. The unit of content was a two-sentence narrative of 30 words, dealing with the earthquake in Alaska in the spring of 1964. CWR values of .83 (25 content words out of the total of 30 words), .67 (20 content words), .50 (15 content words), and .33 (10 content words) were studied.

Materials consisted of 32 units of content—eight for each value of CWR. Each unit of content of an eight-unit series represented a subtopic of the general topic—the Alaska earthquake. Each of the eight subtopics was written four ways, so as to reflect each of the CWR values. A loss of detail necessarily occurred as versions of a subtopic were varied from high to low CWR. Subtopics were ordered the same way for each of the four CWR series.

The following illustrate the materials used:

CWR=.83. News of the recent serious quake in Southern Alaska reached West Texas residents Saturday morning. Early wire service reports indicated it caused great water and shock damage in downtown Anchorage.

CWR=.67. Little news concerning the quake in Alaska reached West Texas residents before Sunday morning. However, wire service reports received Saturday night indicated that it was severe and damage quite extensive.

CWR=.50. Little news concerning the quake in Alaska reached Texas before Sunday morning. However, enough information was available by Saturday night to indicate that its effects were severe and quite extensive.

CWR=.33. Very little news concerning the quake in Alaska had been received by Sunday. However, at least some commentators were reporting on Saturday that not very many people had been killed.

Apparatus. Units of content were presented on slides by means of rear-view projection onto a viewing surface. Slide presentation and response intervals for each exposure were timed by specially designed equipment and times printed out.

Subjects. Sixty-four first-year Army enlisted men—high school graduates aged 18 to 23 years—served as subjects. They were selected from a pool of first-year men on the basis of ability to follow instructions and freedom from marked reading problems.

Procedure. Sixteen men were assigned to each treatment, according to the order in which they arrived at the laboratory. Each subject, seated before the screen in a small room, was told that a series of slides, each containing a two-sentence paragraph, would be presented on the screen. He was instructed to read the material one time carefully; press a button at his right, which would make the screen go blank; report the material contained on the slide, using only the words in which the two sentences were written; and then press the button a second time. It was emphasized that any changes in word order that did not change the meaning of a sentence would be acceptable. Finally, the subject was told that the same slide would be presented repeatedly until he could recite all of the material on the slide without error.

The interval during which the slide was in view for a given exposure was its presentation time, and the interval between the subject's first and second button presses was the response time for that exposure. The measures used in analyzing the data were cumulative presentation time and cumulative response time, over exposures, to a criterion of one perfect immediate recall of the unit of content.

Before the eight-slide experimental series was presented, the subject learned two paragraphs of the same form to a criterion of one perfect immediate recall. These practice materials were topically unrelated to the experimental materials.

RESULTS

Treatment means and standard deviations for cumulative presentation time and response time to criterion are presented in Table 1. F ratios ($df = 3, 60$) for presentation time and response time data were 2.94 and 3.41,

Table 1

Presentation and Response Times: Study 1
(seconds)

Content Word Ratio (CWR)	Presentation Time		Response Time	
	Mean	Standard Deviation	Mean	Standard Deviation
.83	183.7	60.5	112.9	55.4
.67	137.9	49.4	81.5	59.8
.50	149.7	66.2	82.2	36.5
.33	123.2	63.2	60.0	30.3

values that are significant at the .05 level. There were significant differences[1] at the .05 level between CWRs of .83 and .67 and between CWRs of .83 and .33 for presentation time, and between CWRs of .83 and .33 for response time. Thus, lowering the CWR below .67 had no statistically significant effect upon either presentation time or response time.

DISCUSSION

In a sample of 79,390 words of telephone conversation, French, Carter, and Koenig (24) found that 65% of the words were nouns, adjectives, adverbs, verbs, or pronouns—words classified as content words in the present study. Similarly, Fries (19) found that in some 50 hours of telephone conversation, approximately 67% of the material consisted of content words.

For a large sample of written English, Miller and Chomsky (25) found that content words comprised only 41% of all words. Pronouns and a few other words are categorized as function words in that study; if these words were removed from the function word class, the CWR of the sample would rise to .49—still considerably lower than the CWR for conversation.

A preliminary analysis of paragraphs taken from Army Field Manuals, and of paragraphs taken from the Background, Procedure, and Discussion sections of this chapter shows a tendency for CWRs to cluster closely around a value of .55. It seems likely that the normal range of CWRs for written material is on the order of .50 to .60.

The results of the present study indicate that use of a Content Word Ratio well above the norms reported, for either written or spoken English, will significantly increase the time required to learn topically related pairs of sentences of varied syntactic form. However, CWR variations within or below the normal range apparently have negligible effect upon learning time.

If a practical conclusion can be drawn from Study 1, it is that discourse written at a fairly high CWR level—at, or just beyond the top of the normal

[1]Duncan's Multiple Range Test (Edwards, 23).

range—represents a relatively efficient form of communication in terms of the amount of information acquired per unit of time.

RELATION BETWEEN CWR AND LENGTH MEASURES

CWR=f (Length in Syllables). The number of syllables in each paragraph was counted, and the length in syllables was then plotted against the mean CWR. The results indicated that CWR and paragraph length in syllables (Syl) tend to be linearly related; for Study 1 materials the relationship between CWR and length in syllables is approximately that shown in the following equation: $CWR = .035Syl - 1.333$.

The reason for the relationship is not difficult to understand. With content words drawn at random from Pool A and function words from Pool B, if there is a pronounced tendency for the words of one pool to be longer than those of the other pool, then—on the average—CWR should be a linear function of number of syllables in the paragraph. There is indeed a pronounced tendency for content words to be longer than function words.

CWR=f (Length in Kernel Sentences). Paragraph length may also be defined in terms of kernel sentences. Kernel sentences are one-clause declarative sentences which, in transformation grammars, are considered elemental constructions in the sense that any further reduction produces non-sentences. Transformation grammars treat ordinary grammatical sentences as manifestations of one or more kernel sentences that have been transformed by the application of one or more rules. For example, a sentence of the form, "Reports indicated that the quake was severe," results from the application of certain transformation rules to two kernel sentences, "Reports indicate X" and "The quake is severe."

As yet, transformation grammars are not complete enough to permit an entirely reliable analysis of written material into its underlying kernels. However, Coleman[1] has developed a method which, by counting certain content words in a passage, yields a reliable measure that seems, intuitively, to be closely related to the average counts of kernel sentences one would obtain using Harris's (**26, 27**) list of kernel sentence types. The Coleman count for approximating the number of kernel sentences in a passage is based on the count of (a) all verbs, including gerunds, infinitives, and nominalizations; (b) all modifiers except those of "N is A" sentences (e.g., "severe" is not counted in "The storm is severe") and noun adjuncts that modify nominalizations (e.g., "shock" is not counted in "shock damage"); (c) all prepositions except those whose objects are nominalized verbs and those in prepositional phrases that modify nominalized verbs; (d) ellipses of all verbs and modifiers.

Coleman applied his system to the 32 paragraphs used in Study 1. His analysis yielded a count for each paragraph that is an estimate of the number of kernel sentences contained in the paragraph. Paragraph length in estimated number of kernel sentences (K) was then plotted against CWR. The relation between CWR and K for the Study 1 materials was found to be linear in form and can be represented by the equation: $CWR = .0444K - .104$.

Cross-Study Comparison of Findings. The measure of load, CWR, and the measures of length, Syl and K, are linearly related for Study 1 materials. Thus, findings based on either measure of length should have the same form as those

[1]Personal communication from Edmund B. Coleman.

based upon CWR. Study 1 results therefore could be accounted for as readily on the basis of paragraph length as on CWR.

The Study 1 data do not support the view that difficulty is an increasing function of the ratio of content words to total words. Consequently, they do not support the view that difficulty is an increasing function of paragraph length. On the other hand, subsequent findings of Studies 2, 3, and 4 (to be reported in the following chapters) dealing with effects of length upon recall, do support the view that difficulty is an increasing function of length. While Study 1 differed in a number of ways from the other three studies, there is no ready explanation as to why the Study 1 findings—considered as a function of length in light of the linear relations demonstrated above—should have deviated from those of the other studies. The discrepancy across studies suggests that Study 1 should be replicated.

Chapter 3

SENTENCE RECALL AS A FUNCTION OF NUMBER OF SENTENCES TO BE RECALLED (Study 2)

Abstract of Study 2

The main objective of the two experiments in this study was to examine the effects of amount of material upon sentence recall. In the first experiment, the effects of the length of the sentence list to be memorized were confounded with type-token ratio, a ratio reflecting amount of word repetition. The effects of list length were evaluated independently in the second experiment.

Whatever the measure, the rate of increasing difficulty with increased amount of material was greater for material of the first experiment than for materials of the second.

BACKGROUND

The effects of list length upon the recall of lists of nonsense syllables vary with other germane characteristics of lists, such as association values of items and their meaningfulness. We would expect the same consequences when the lists consisted of sentences, although the range of germane characteristics of lists of sentences is less well understood at present than is the range of characteristics pertinent to studies involving nonsense syllable lists.

The main objective of the two experiments in this study was to examine the effects of list length in sentences upon sentence recall, while controlling for certain apparently-germane associated characteristics of the materials. If this approach proved fruitful, then the list format might lend itself to the analytic study of various characteristics of written material.

Besides tests of difference between conditions, techniques for examining effects included empirical curve fitting to determine the relationship—expressed in mathematical terms—between independent and dependent variables. Such curves may eventually provide the basis for mathematical equations to represent the relationship between learning and the linguistic characteristics of technical written material.

METHOD

Materials. In Experiment 2a units of content were lists of two, three, four, or five sentences (S). Each sentence contained one clause that expressed a relation holding between three one-syllable, five-letter surnames—"N_1 V N_2 to N_3"—for example, "Locke acclaimed Chase to Hodge." The sentences of a slide were constructed from a vocabulary of six surnames and one verb in such a way that no syntactic overlap occurred among surnames from sentence to sentence of the set contained on the slide.[1] Thus, no two of the surnames appeared at N_1, no two at N_2, and no two at N_3.

[1]When two sentences have the same subject or the same object, or share certain other syntactic functions, there is syntactic overlap between them. There is syntactic overlap between "Bill hit the ball" and "Bill chased the dog," since these two sentences could be rewritten as a single sentence with a compound predicate—"Bill hit the ball and chased the dog."

Seven content word types (six surnames, one verb) were used on each slide; these were used to provide the token content words (nouns, verbs) needed for each set of sentences. Thus, there were four times as many content words as sentences on the slide. Consequently, while treatments (S = 2, 3, 4, or 5) were devoid of syntactic overlap, they varied in Type-Token Ratio (TTR) for content words as follows: (a) S=2, TTR=.87, (b) S=3, TTR=.58, (c) S=4, TTR=.44, (d) S=5, TTR=.35. There were no sequential constraints on the set of sentences. Thus, any order of recall of sentences in a set was considered to be correct.

The following two- and four-sentence sets are illustrative of the materials used in Experiment 2a:

S=2. Trask assigned Lange to Wells.
Simms assigned Price to Cloyd.
S=4. Wertz described Brock to Suggs.
Brock described Young to Greer.
Downs described Suggs to Wertz.
Young described Wertz to Downs.

In Experiment 2b the materials consisted of sets of one, two, three, four, or five one-clause sentences (S). Each sentence expressed a relation between a two-syllable noun used as subject and a two-syllable noun used as object—"N_1 V N_2"—for example, "The builder disliked the youngster." Verbs also were two-syllable. The mean frequency in print of the content words (nouns, verbs) used on a slide was approximately 20 per million (G-count, Thorndike and Lorge, 18). The sentences of a slide were constructed with twice as many different nouns as sentences, and the same number of different verbs as sentences. Thus, unlike the materials of Experiment 2a, those of 2b repeated no content words across a set of sentences—that is, the ratio of type to token content words was one, in every case. There were no sequential constraints on the set of sentences; all orders of recall were considered correct.

The following are illustrative of the materials used in Experiment 2b:

S=1. A scholar annoyed the tourist.
S=3. A surgeon devised the charter.
A dealer procured the license.
A grocer landscaped the hillside.
S=5. The trainer cautioned the pitcher.
The barber restrained the batter.
The runner dismayed the waiter.
The actress outlined the complaint.
The miner adjourned the hearing.

Apparatus. The apparatus was the same as that used in Study 1, except that during Experiment 2a intervals were timed by electric clocks and the data recorded manually.

Design. Treatments-by-subjects designs were used in both experiments. In Experiment 2a, subjects received a practice series consisting of four slides, followed immediately by an experimental series of 12 slides. One each of the practice series and three each of the experimental series contained two, three, four, and five sentences of form "N_1 V N_2 to N_3." The three slides representing each treatment in the experimental series (S=2, 3, 4, 5) were presented in series. These triples were counterbalanced for order.

In Experiment 2b, subjects received a practice series consisting of five slides, followed immediately by an experimental series of 20 slides. One each of the practice series and four each of the experimental series contained

one, two, three, four, and five sentences of form "N_1 V N_2." The four slides representing each treatment in the experimental series (S = 1, 2, 3, 4, 5) were presented in series. These quadruples were counterbalanced for order.

The following data were obtained: (a) cumulative presentation time and cumulative response time over exposures to a criterion of one perfect immediate recall of the unit of content and (b) "trials" (exposures) to criterion.

Subjects. The eight subjects of Experiment 2a were volunteers from a pool of HumRRO civilians and military research assistants. Subjects ranged in age from 25 to 41 and in education from two to seven years of college. The 16 subjects of Experiment 2b were first-year Army enlisted men who were high school graduates and 18 to 23 years of age. Subjects had been screened as in Study 1 on the basis of ability to follow instructions and freedom from marked reading problems.

Procedure. Procedure followed that for Study 1 except that no deviation from word-order within the sentence was allowed.

RESULTS

Statistical Analyses

Treatment means and standard deviations for cumulative presentation time, response time, and total time, and for number of "trials" to criterion, by experiment, are presented in Table 2. For all measures in both studies, differences between treatment means[1] are significant at the .05 level.

Empirical Fits

That increasing the amount of material should significantly increase the difficulty of recall is virtually a truism. The main value of the Study 2 results

Table 2

Presentation, Response, and Total Times, and "Trials" to Criterion: Experiments 2a and 2b

Number of Sentences	Experiment 2a				Experiment 2b			
	Presentation Time (sec)	Response Time (sec)	Total Time (sec)	Number of "Trials"	Presentation Time (sec)	Response Time (sec)	Total Time (sec)	Number of "Trials"
1 S								
Mean	–	–	–	–	3.6	2.9	6.5	1.04
SD	–	–	–	–	1.8	.8	2.0	.08
2 S								
Mean	25.9	21.5	47.4	2.14	12.3	8.1	20.4	1.30
SD	8.4	8.5	13.1	.76	7.2	4.1	9.3	.36
3 S								
Mean	65.4	49.7	115.2	2.62	38.4	22.8	61.2	1.87
SD	17.6	22.4	26.5	.76	16.6	9.7	21.0	.43
4 S								
Mean	123.1	105.0	228.1	4.05	70.0	41.6	111.6	2.29
SD	10.2	64.0	64.3	2.25	36.1	30.7	58.7	1.02
5 S								
Mean	261.6	229.6	491.2	7.82	92.4	63.6	156.1	2.91
SD	118.0	124.7	212.3	5.41	32.0	29.3	49.1	.93

[1]Evaluated using Friedman's Two-way Analysis of Variance by Ranks (Siegel, 28, Chapter 7, pp. 166-172).

rests with the form of functional relationship holding between measures of recall and amount of material.

The Equations Used. The equations most frequently used to relate independent variables (X) to dependent variables (Y) in the studies in this report are of three types[1] ("e" represents the base of the natural logarithm; other lower case letters represent empirical constants to be fitted from data):

Type 1: $Y = aX^n$

Type 2: $Y = ae^{bX}$

Type 3: $Y = ae^{bX^n}$

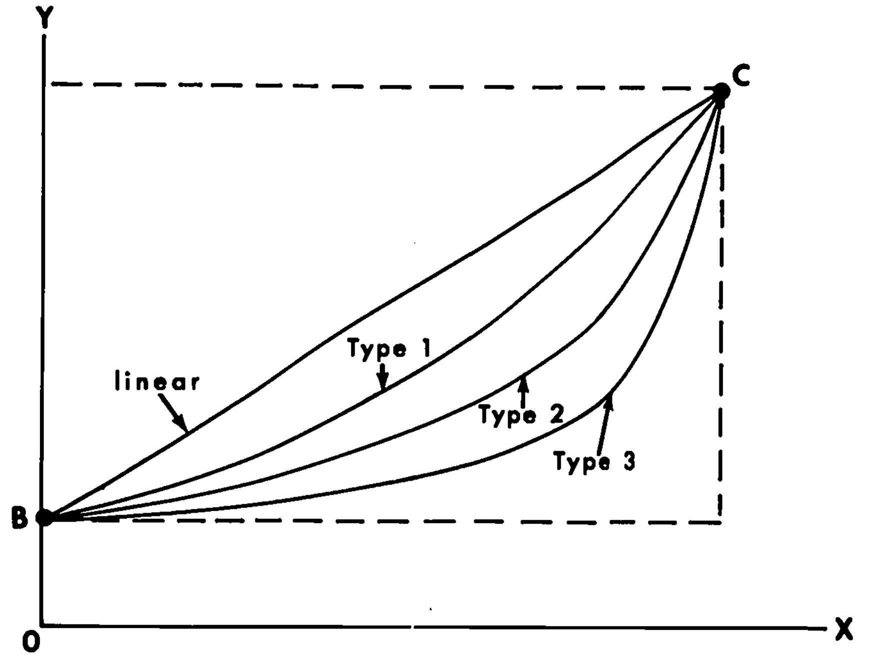
Where n is positive and greater than one for the Type 1 equation, b is positive for the Type 2, and b and n positive and n greater than one for the Type 3, these equations represent functions that are monotonic, increasing, and positively accelerated. That the equations belong to a single "family" can be demonstrated by rewriting them to a common base of Y as equal to a function of X represented as a power of 10:

$Y = aX^n$ becomes $Y = a10^{n(\log X)}$

$Y = ae^{bX}$ becomes $Y = a10^{sX}$

$Y = ae^{bX^n}$ becomes $Y = a10^{sX^n}$

Equations of the three types may be represented visually with respect to "bend"—that is, in terms of deviation from a straight line relationship. In the accompanying diagram, the area under the "diagonal" is the domain of such equations when their functions are positively accelerated and increasing.



The degree of "bend" depends on the form of a function's acceleration, differing from one type of equation to another. More "bend" can be obtained

[1]Lewis (29) calls mononomials of the form $Y = aX^n$ parabolic-type equations. The Type 2 equation is a simple exponential equation. There is no conventional name for the Type 3 equation. On the basis of the fact that its parameter X is exponented and that its basic form is exponential, the Type 3 equation can be described as a parameter-exponented exponential equation.

from a Type 3 equation than from a Type 2, more from a Type 2 than from a Type 1. The degree of "bend" in a function reflects the degree of increasing difficulty in recalling material as the amount of material increases.

All empirical fits of the data were obtained using the least squares method—Lewis (29)—except that for the Type 3 equation, which was fit according to a mixed graphic-least squares procedure described in Chapter 5.

Fits for the Time Data. Empirical fits of the Experiment 2a and Experiment 2b means for total time—presentation time plus response time—are presented in Figure 1. The curves portray the effects of number of sentences to be recalled (S) on mean total time (Y). The symbol Y' represents computed (predicted) Y.

Fits that appeared by inspection to be reasonably good were obtained for presentation time means and for response time means using equations of the same form as those used to fit total time means. An interesting aspect of the data for the two experiments is that Type 2 equations describe functions of S for the Experiment 2a time measures, while Type 1 equations describe functions of S for the Experiment 2b time measures.

One may ask what happens to mean total time per sentence as the number of sentences to be recalled increases. That is, what is the nature of $Y'/S = f(S)$? On the basis of the equation derived from the empirical fit of the Experiment 2a total time means, the Y'/S values for $S = 1$ to 5 are approximately 23, 25, 36, 58, and 101 seconds, respectively. Time per sentence is a Type 3 function of number of sentences.

Since Y' is approximately equal to aS^2 for the Experiment 2b total time data, Y'/S tends to increase linearly with increasing S. Thus, for Experiment 2a materials and conditions, total time per sentence is an extremely positively accelerated function of number of sentences to be recalled, whereas, for Experiment 2b materials and conditions, the difficulty of recall increases linearly with increasing S and acceleration is zero.

Fits for the "Trials" Data. Mean "trials" (exposures) to criterion is not the same function of S as are means for the time measures. For both experiments, mean "trials" is a more sharply accelerated function of the number of sentences to be recalled than is mean time. For Experiment 2a "trials" to criterion means, a good fit was obtained using a Type 3 equation—$Y' = 1.86e^{.018S^{2.8}}$. For Experiment 2b mean "trials" data, a fair fit was obtained using a Type 2 equation—$Y' = .82e^{.26S}$.

Whether for time or "trials" measures, functions that describe the Experiment 2a data manifest more "bend"—accelerate more sharply—than functions describing the Experiment 2b data. This is also the case for measures of the Y'/S type—as in total time per sentence. The pattern of increasing difficulty with increasing S is one of a faster rate of increasing difficulty for the Experiment 2a data than for the Experiment 2b data.

Time Per "Trial" Functions. The ordinal values of Figure 2 were obtained by dividing the fitted values of mean total time for each value of S in each experiment by mean "trials" to criterion for that value of S. This quotient represents average total time per "trial."

It is evident that the plot of this quotient against S is linear for the Experiment 2b data. For the Experiment 2a data, the plot falls on a line for $S = 2, 3, 4$, but departs from linearity when $S = 5$ is considered. The average amount of total time used by subjects over t "trials" to criterion tends to be linearly related to the number of sentences that must be read and reported. On the average, then, when presentation time is controlled by the subject, he acts

Effects of Number of Sentences To Be Recalled on Total Time:
Fits of Experiments 2a and 2b Means

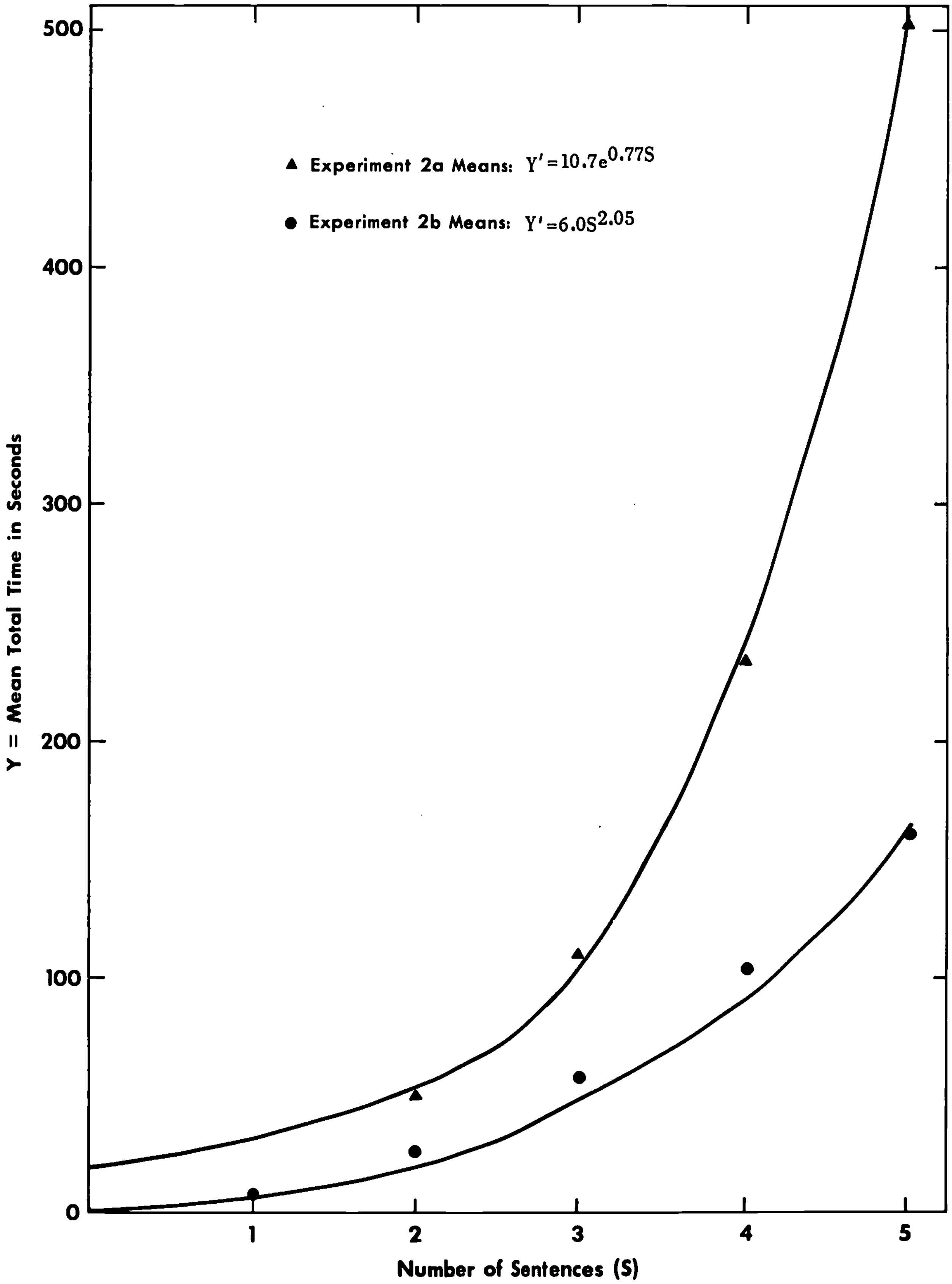


Figure 1

Effects of Number of Sentences To Be Recalled on Mean Predicted Total Time Per Mean Trial: Experiments 2a and 2b

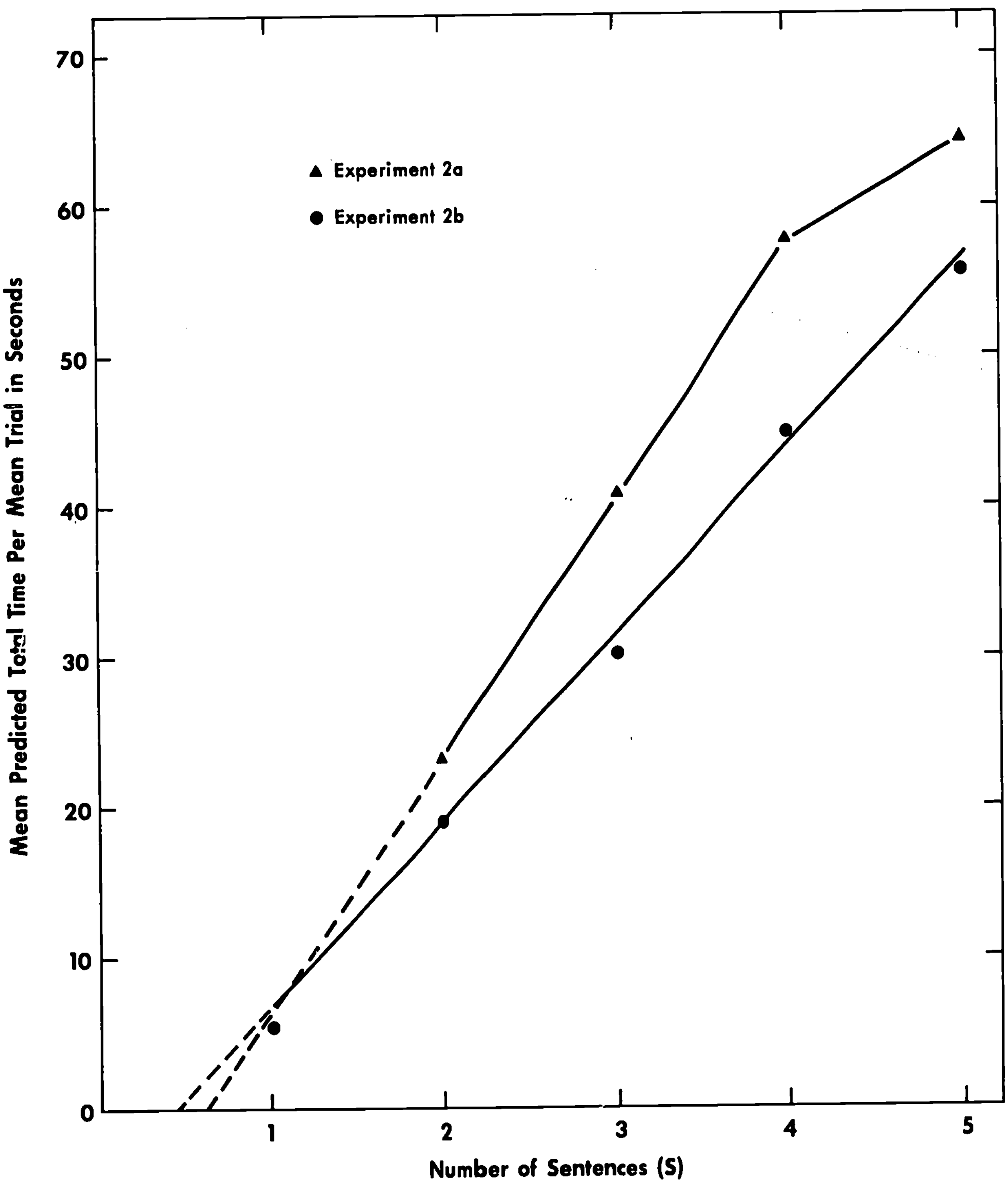


Figure 2

as if he should give one unit of time to each unit of material—sentence—during a "trial." (There is a good deal of variation around this average performance.)

DISCUSSION

A number of factors varied from Experiment 2a to Experiment 2b, so it is not possible to account with certainty for the difference in form of functions of

S between these experiments. However, it is clear that the rate of increasing difficulty of recall with increasing amount of material is considerably greater for the Experiment 2a materials than for those of 2b. We would guess that such differences stem primarily from letting the Experiment 2a TTRs take values less than unity under the condition of no syntactic overlap.

There is a point of view in readability research that lowering the ratio of type to token words below unity will ease recall. When the number of different words in 100 words of running prose—a TTR—is used as a factor in a readability formula, the factor is weighted to yield a readability score which predicts that readability will improve as the number of different words per 100 words of running prose decreases. While this might be the case when repeated words repeat the syntactic functions of their earlier usage—that is, when repetition is redundancy—the present results lead us to suspect that the opposite may be true when word repetition is not accompanied by repetition of syntactic function. That is, word repetition without syntactic overlap might hinder rather than facilitate readability.

The Experiment 2b data are compared with portions of the Study 4 data in Chapter 5.

Chapter 4

RECALL OF SETS OF SENTENCES AS A FUNCTION OF CONTENT WORD LENGTH (Study 3)

Abstract of Study 3

Two experiments were conducted to determine the effects of word length, defined as number of syllables in the word, on learning a set of sentences. The recall of a set of three "Noun is adjective" sentences as a function of mean length, in syllables, of the six content words in the set was tested. In one experiment, the pattern of syllabic length across the words of a sentence was the same from sentence to sentence of the three-sentence set; in the other, the syllabic pattern varied from sentence to sentence.

In both studies, time and "trials" measures of recall showed increases with increasing length of content words in the set of sentences to be recalled.

BACKGROUND

The use of the percent of polysyllabic words as an indicator of reading difficulty was first suggested by G.R. Johnson (30) in 1930. Since then, word length measures have been widely employed in readability formulas. In spite of this, the evidence is meager that word length per se is related to the immediate recall of written material.

Zipf (31) demonstrated in 1935 that longer words tend to occur less frequently in the language system. It has since been shown that recall is greater for words with a high English text frequency than for words with a low frequency (Hall, 32; Sumby, 33; Lloyd, 34). Hence, we would expect longer words to be more difficult to recognize and to recall on familiarity grounds alone. Word length has, however, been shown to be a significant determinant of the word recognition threshold when word frequency effects have been partialed out (McGinnies, Comer, and Lacey, 35; Newbigging and Hay, 36). Word duration thresholds are a linear, increasing function of word length.

The purpose of the present experiments was to determine whether word length, defined as number of syllables in the word, has an appreciable effect on learning a set of sentences of identical syntactic form, where vocabularies are matched for frequency in print.

METHOD

Materials. In both experiments, the unit of content was a set of three sentences of form "N is A"—for example, "The writer was hungry." In this particular sentence, the noun contains two syllables and the adjective two syllables, so the sentence pattern (SP) is denoted 22. Each of the three-sentence sets of Experiment 3a was homogeneous for sentence pattern; thus, the three sentences of a set might have the SPs 21, 21, 21. Those of Experiment 3b were heterogeneous for SP; the three sentences of a set might have the SPs 21, 11, 22.

In both of these illustrations, the mean Content Word Length (CWL) is 1.5 syllables. Both experiments varied the CWL for the six content words appearing in a set of three sentences.

In Experiment 3a, the CWL values were 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 syllables. These CWL values were based, respectively, on SPs of 11, 21, 22, 32, 33, 43, and 44.

A practice series consisting of seven slides (one for each value of CWL) and an experimental series consisting of 28 slides (four for each value of CWL) were used in Experiment 3a. G-count frequencies (Thorndike and Lorge, 18) for nouns and adjectives used in experimental slides ranged from nine to 40 per million words. Mean frequency was 20 per million. No semantically anomalous sentences—for example, "The shovel was thirsty"—were used. Sentences were arranged in list format on the slide.

The following illustrate Experiment 3a materials:

CWL = 1.0. The dam was crude. The ranch was neat. The creek was quaint.
CWL = 2.0. The mayor was stubborn. The grocer was fearless. The fighter was clever.
CWL = 3.0. The producer was efficient. The attendant was picturesque. The conductor was sensible.
CWL = 4.0. The reservation was mysterious. The missionary was victorious. The economy was astonishing.

Excepting that sentences of a slide were heterogeneous for SP, the materials of Experiment 3b were the same as those of 3a. The following example illustrates SP heterogeneity: (a) The first sentence of a set contains a two-syllable noun and a three-syllable adjective (SP = 23), (b) the second sentence a four-syllable noun and a one-syllable adjective (SP = 41), and (c) the third sentence a three-syllable noun and a two-syllable adjective (SP = 32). The CWL values of Experiment 3b were 1.5, 2.0, 2.5, 3.0, and 3.5 syllables. Materials consisted of 25 slides—one practice and four experimental slides for each value of CWL.

The following illustrate Experiment 3b materials:

CWL = 1.5. The convict was deaf. The tramp was bruised. The builder was wealthy.
CWL = 3.5. The management was courteous. The brotherhood was indifferent. The legislation was unusual.

Apparatus. The apparatus was the same as used in Study 1.

Design. A treatments-by-subjects design was used in both experiments. The same subjects were used, half participating in Experiment 3a first and half in 3b first. The following data were obtained: (a) cumulative presentation time and cumulative response time over exposures to a criterion of one perfect immediate recall of the unit of content, and (b) "trials" (exposures) to criterion.

Subjects. Subjects were 16 first-year Army enlisted men who had participated in Study 1.

Procedure. Procedure followed that for Study 1 except that no deviation from word-order within a sentence was allowed. In both experiments the practice slides were presented in order of ascending CWL. The four slides representing each treatment value of an experiment were grouped in the series for that experiment. These treatment quadruples were randomly ordered in the experimental series.

RESULTS

Treatment means and standard deviations for cumulative presentation time and response time and for "trials" to criterion, by experiment, are presented in Table 3. Treatment results differed significantly on all measures. Experiment 3a F ratios ($df = 6, 90$) for presentation time, response time, and "trials" to criterion are, respectively, 29.4, 15.8, and 5.83, values that are significant

Table 3

Presentation and Response Times, and "Trials" to Criterion: Experiments 3a and 3b

Content Word Length (CWL)	Experiment 3a			Experiment 3b		
	Presentation Time (sec)	Response Time (sec)	Number of "Trials"	Presentation Time (sec)	Response Time (sec)	Number of "Trials"
1.0						
Mean	16.9	7.6	1.32	–	–	–
SD	4.9	3.6	.36	–	–	–
1.5						
Mean	19.9	9.0	1.33	14.1	8.0	1.27
SD	5.5	3.0	.25	4.7	3.1	.30
2.0						
Mean	22.3	10.1	1.43	19.0	9.7	1.33
SD	9.1	5.8	.49	6.3	3.[illegible]	.26
2.5						
Mean	23.2	11.6	1.58	26.1	14.3	1.82
SD	6.7	3.9	.35	8.4	5.1	.37
3.0						
Mean	40.2	19.0	1.96	29.2	14.6	1.61
SD	14.0	7.8	.62	12.0	8.3	.62
3.5						
Mean	47.2	18.7	1.80	32.2	16.2	1.71
SD	23.0	9.2	.80	12.5	8.5	.73
4.0						
Mean	47.3	17.9	1.81	–	–	–
SD	17.8	7.0	.57	–	–	–

at the .005 level. Experiment 3b $\underline{F}$ ratios ($\underline{df}=4, 60$) for the same measures are 23.8, 6.11, and 4.24, also significant at the .005 level.

For the presentation time data of Experiment 3a, only the means of three adjacent pairs (CWL = 1.5–2.0, 2.0–2.5, and 3.5–4.0) failed to differ significantly at the .05 level.[1] The same trend is apparent for the Experiment 3b presentation time data, wherein only the means of two adjacent pairs (CWL = 2.5–3.0 and 3.0–3.5) failed to differ significantly.

For the response time and "trials" data of both experiments, the CWL values below 2.5 and above 2.5 tended to form groups such that the means of no two pairs within a group differed significantly; however, the means of all between-groups pairs differed significantly at the .05 level.

Means for presentation and response time measures for the two experiments are plotted against CWL in Figure 3. An inspection of these plots—particularly for the Experiment 3a measures—shows scant support for the view that the function Y = f (CWL) is monotonic. However, the plots are consistent with the proposition that the form of the function of Content Word Length will not be the same when materials are homogeneous for sentence pattern as when they are heterogeneous for sentence pattern.

[1]The Wilcoxin Matched-Pairs Signed-Ranks Test (Siegel, 28, pp. 75-83) was used to find differences between treatment pairs for all measures.

Effects of Content Word Length on Presentation and Response Times: Experiments 3a and 3b

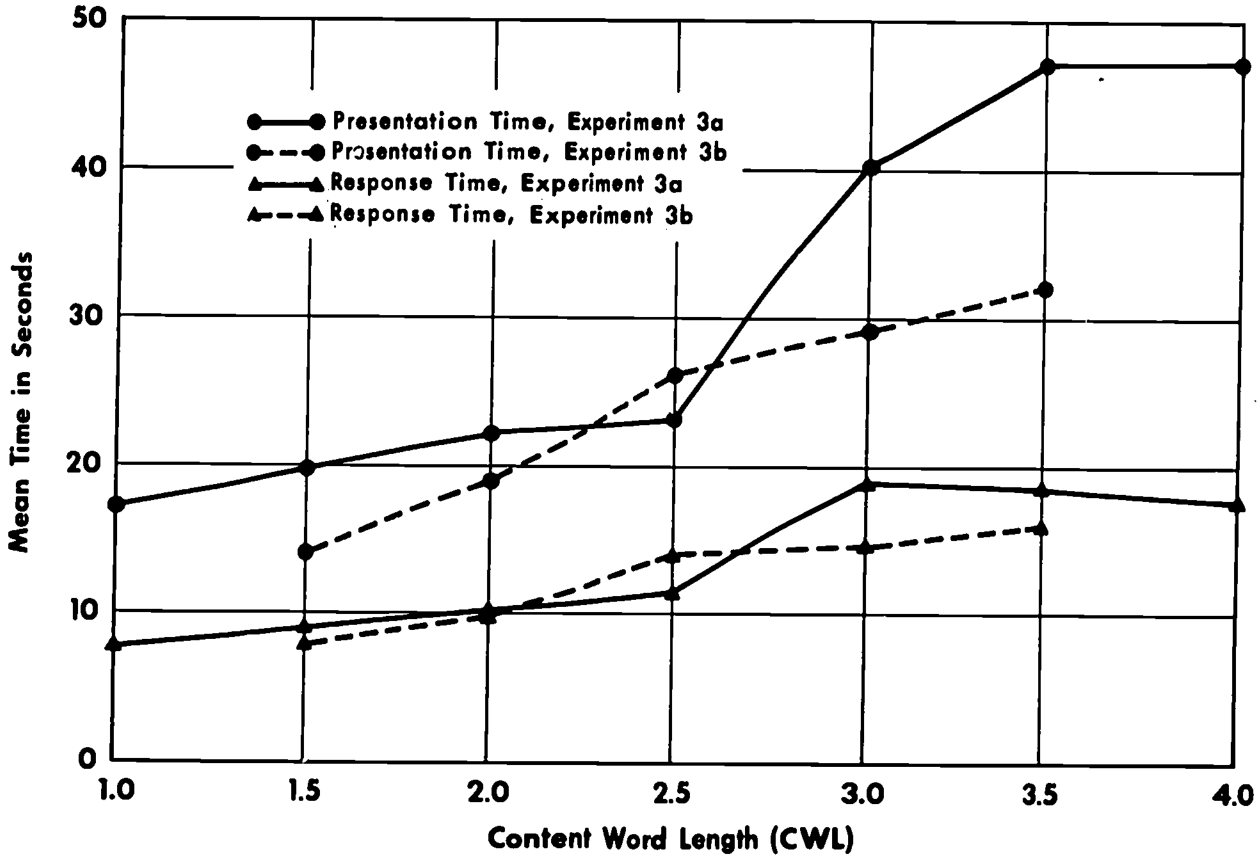


Figure 3

DISCUSSION

The present results indicate that word length has a considerable effect upon the time required to learn a list of unconnected sentences. Whether increments are as substantial when sentences are topically related and/or of varied syntactic form remains to be determined. However, it seems unlikely that the word length effect is peculiar to the particular sentence form used in these studies. (The difference in shape of function between experiments—suggested in the Figure 3 plots—raises the possibility that shape of the function may be sensitive to all sorts of systematic variations in the material.)

It should be noted that word length may be defined in a number of ways. In visual and auditory threshold studies, word length has been measured in letters, in phonemes, and in syllables. Readability formulas have most commonly dealt with word length in syllables or, infrequently, in terms of affixed morphemes—prefixes, suffixes, and infixes of the class of function morphemes.

The choice of a particular measure may not be of much practical importance, since all these measures are highly although imperfectly correlated. For the nouns and adjectives used in Experiment 3a, for example, the correlation between number of letters and number of syllables was +.87; for nouns and adjectives of Experiment 3b, it was +.83. There was an inverse relation between syllabic word length and mean number of letters per syllable. This relation may account for the fact that increments in learning time were slight as mean content word length increased beyond three syllables.

Chapter 5

SENTENCE RECALL AS A JOINT FUNCTION OF NUMBER OF KERNELS IN THE CLAUSE AND NUMBER OF CLAUSES IN THE SENTENCE

(Study 4)

Abstract of Study 4

In exploratory pilot research based on a single subject, the number of kernel sentences (K) in the clause and the number of clauses (C) in the sentence were systematically varied. Within the limits of the data base, mean presentation time to criterion was adequately described by an equation of the form $Y' = ae^{bC g^{K}}$. That is, mean presentation time to criterion was a jointly exponential function of number of clauses and number of sentences in the clauses.

BACKGROUND

Recent grammars tend to be in two parts: (a) a surface grammar dealing with structural characteristics of the language as it is used and (b) a deep grammar dealing with elemental constructions—kernel sentences—and transformation rules that relate structures of the deep grammar to those of the surface grammar.

The clauses (C) of this study are analogous to the sentences (S) of Experiments 2a and 2b. Experiments 2a and 2b demonstrated that when all clauses are simple and of the same form, then any one of a number of measures of recall will be a robust function of number of clauses in a list of coordinate clauses. Will the number of kernel sentences (K) in the clause have as profound an effect upon a measure of recall as the number of clauses (C) in the sentence? Answering this question can be approached by answering the broader question, "What joint function of C and K is sentence recall?" A preliminary evaluation was undertaken in pilot research employing a single but well-practiced subject.

METHOD

Materials

Kernel Sentences. Different grammars employ the concept of the kernel sentence somewhat differently. They vary (a) in their inventories of types of kernel sentences and (b) in the kernel sentences that, according to their rules, will be extracted from a given sentence in normal form. This need not be of concern here, since the kernel sentence types used in these experiments form a quite restricted set.

Perhaps there would be general agreement that a clause of form "N_1 P A_1 N_2 is A_2"—for example, "A class on good plays would be small"—consists of three kernel sentences: (a) "N_1 is A_2"—"Class is small," (b) "N_1 is P N_2"—"Class is on plays," and (c) "N_2 is A_1"—"Plays are gooc " This is the extent to which the present research has needed to cope with the problem of analyzing clauses into their kernel sentences.

The three kernel sentences of the foregoing example are of two types: (a) "N is A" and (b) "N is P N". On the basis of crude pilot studies antedating

the present research, it would appear that not all kernel sentence types have the same effects upon recall. However, the earlier research supports the notion that these two types are informationally equivalent in the sense that they have the same effects on recall. Therefore, it is assumed that types "N is A" and "N is P N" are interchangeable as units of clause length.

K x C Sentences. The content units of the present research represent all combinations of K = 1, 2, 3 and C = 1, 2, 3. In Experiment 4a, the clauses of a multiple-clause sentence (C = 2, 3) were joined by and. In Experiment 4b, the clauses of such sentences were in an if-then relation. The sentence forms of the two studies are shown in Table 4.

Table 4

Sentence Forms of Experiments 4a and 4b

Experiment	Kernel Sentences	Clauses[a]		
		C=1	C=2	C=3
4a	K=1	N is A	N is A and N A	N is A, N A, and N A
	K=2	N P N is A	N P N is A and N P N A	N P N is A, N P N A, and N P N A
	K=3	N P A N is A	N P A N is A and N P A N A	N P A N is A, N P A N A, and N P A N A
4b	K=1	N is A	N is A if N is A	If N is A, then N is A if N is A
	K=2	N P N is A	N P N is A if N P N is A	If N P N is A, then N P N is A if N P N is A
	K=3	N P A N is A	N P A N is A if N P A N is A	If N P A N is A, then N P A N is A if N P A N is A

[a]N = Noun; A = Adjective; P = Preposition.

If "would be" is substituted for "is" and an article—"a"—is counted at the head of each clause, then Table 4 can be used to count total words in the sentence, function words in the sentence, and content words in the sentence. All words were one-syllable, so sentence length in syllables equals number of words in the sentence. All words used in the materials of these experiments appear in Part I of the G-count of Thorndike and Lorge (18).

Illustrative Materials. The following illustrate the materials used in Experiment 4a. The materials used in Experiment 4b were the same as those of 4a except that clauses stood in an if-then relation, as indicated in Table 4.

C = 1, K = 1. A duck would be large.
C = 1, K = 2. A loaf of bread would be fresh.
C = 1, K = 3. A rack for new guns would be tall.

C = 2, K = 1. A dog would be cold and a pig sly.
C = 2, K = 2. A brace of pots would be dear and a school of sharks rare.
C = 2, K = 3. A sack of fine grain would be drab and a load of smooth poles cheap.

C = 3, K = 1. A rat would be brown, a whale plump, and a stag clean.
C = 3, K = 2. A crate of fruit would be large, a batch of corn clean, and a flask of wine light.
C = 3, K = 3. A class on good plays would be small, a vote on wise laws close, and a rule on hard rains rash.

Procedure

The subject had participated in a series of closely related studies prior to the present research. Data were collected in three sessions. During Session 1, C = 1 sentences were recalled; during Session 2, C = 2 sentences; and during Session 3, C = 3 sentences. Three sentences were recalled under each treatment condition. During a session, nine sentences associated with Experiment 4a and nine associated with 4b were memorized. Ten-minute breaks

occurred between sessions. The first and second sessions consumed very little time. During the third session, the subject was in the experimental situation for approximately 12 minutes, although not responding continuously throughout the period. Total cumulative plus response time during Session 3 was just over nine minutes.

Sentences were presented on 5x8 cards. Otherwise, procedure was the same as that of Study 1.

Subjects

A military research assistant, well-practiced by virtue of participation in several similar studies, was the sole subject.

RESULTS

Presentation Time

Means. Treatment means for cumulative presentation time to criterion, by experiment, are presented in Table 5. The C=1 rows for Experiments 4a and 4b means are the same because the same one-clause sentences occur in the and and the if-then series. Connectives cannot have a role until the sentence contains at least two clauses.

Table 5

Presentation Times: Experiments 4a and 4b
(seconds)

Experiment	Number of Clauses	Number of Kernels K=1	K=2	K=3
4a	C=1	1.51	2.07	2.77
	C=2	2.41	4.25	10.96
	C=3	4.31	11.21	43.88
4b	C=1	1.51	2.07	2.77
	C=2	2.96	6.30	14.81
	C=3	5.92	16.04	49.22

Empirical Fits. $Y' = f(C)$ is in every case a monotonic expression, increasing with positive acceleration. It was found that the data can be represented as a Type 3 equation: $Y' = ae^{bC^n}$

The data were fitted in two stages: (a) First n was evaluated by a graphic method on the basis that log Y is a linear function of C^n. (b) Then a and b were evaluated by the least squares method, with C^n set at values obtained in consequence of graphic evaluation of n.

Empirical fits were as follows:

Experiment 4a: K=1: $Y' = 1.18e^{.25C^{1.48}}$

K=2: $Y' = 1.40e^{.40C^{1.55}}$

K=3: $Y' = .69e^{1.38C^{1.00}}$

Experiment 4b: K=1: $Y' = .76e^{.68C^{1.00}}$

K=2: $Y' = .64e^{1.20C^{.90}}$

K=3: $Y' = .10e^{3.33C^{.57}}$

The parameter K was brought into the equation as a component of the exponent of C. It was assumed that N = gK and that $Y' = ae^{bC^{gK}}$. In consequence, the equations were modified as follows:

$$\text{Experiment 4a: } K=1: \; Y' = 1.18e^{.25C^{1.48K}}$$
$$K=2: \; Y' = 1.40e^{.40C^{.77K}}$$
$$K=3: \; Y' = .69e^{1.38C^{.33K}}$$
$$\text{Experiment 4b: } K=1 \; Y' = .76e^{.68C^{1.00K}}$$
$$K=2: \; Y' = .64e^{1.20C^{.45K}}$$
$$K=3: \; Y' = .10e^{3.33C^{.19K}}$$

The values of the empirically fitted constant b for a condition (for Experiment 4a—.25, .40, 1.38; for Experiment 4b—.68, 1.20, 3.33) form series such that b is a Type 3 function of K. That is, $b = ie^{jK^{1}}$.

It will be recalled that n = gK. A similar series may be noted for g. The values of g for Experiment 4a—1.48, .77, .33—and for Experiment 4b—1.00, .45, .19—form series such that g is a Type 3 function of K, although a negatively accelerated, decreasing one. That is, $g = ue^{-vK^{w}}$.

Thus, the data are described by Type 3 equations having C and K as parameters—$Y' = ae^{bC^{gK}}$—with the constants b and g being functions of K which are of the same form as Y is of C and K. For either experiment, then, the data are consistent with a mathematical surface Y = f(C, K), which can be described by the gaudy expression:

$$Y' = ae^{(ie^{jK^{1}})C^{(ue^{-vK^{w}})K}}.$$

The aspects of the foregoing development that we consider of interest are the following:

(1) For all values of K, cumulative presentation time is a positively accelerated, increasing Type 3 function of C.
(2) The fitted constants n form a linear function of K when, for the equation n = gK, g is a negatively accelerated, decreasing Type 3 function of K.
(3) The fitted constants b form a positively accelerated, increasing Type 3 function of K.
(4) The picture is the same for the materials of both experiments.

Presentation Time Per Word

The mean cost, in seconds of presentation time, of memorizing a word is shown in Table 6 by treatment and experiment.

It is evident that the cost in mean presentation time to a criterion of perfect recall per word (or per syllable) tends to be positively accelerated and increasing for C = 2, 3 and for K = 2, 3, but that the value for K = 1, C = 1 is too high. The conditions of measurement were satisfactory for the longer response intervals associated with two-clause and three-clause sentences, where a slight latency in the experimenter's response to the subject would not affect the relative magnitude of the measure of presentation time to criterion. However, it is conceivable that a small mean constant latency in the experimenter's

response would adversely affect the relation of values for C = 1 sentences to those for C = 2 and C = 3, particularly when the data are of the sort presented in Table 6.

Let us assume that all mean presentation time values of Table 5 are inflated by a half-second. Mean costs per word based upon this assumption are shown in Table 7. Using Table 7 entries as values of mean presentation time

Table 6

Mean Presentation Time Per Word: Experiments 4a and 4b
(seconds)

Experiment	Number of Clauses	Number of Kernels K = 1	K = 2	K = 3
4a	C = 1	.30	.30	.35
	C = 2	.27	.33	.73
	C = 3	.36	.62	2.09
4b	C = 1	.30	.30	.35
	C = 2	.30	.45	.93
	C = 3	.37	.73	1.97

Table 7

Mean Presentation Time Per Word, Adjusted for Estimated Response Latency:[a] Experiments 4a and 4b
(seconds)

Experiment	Number of Clauses	Number of Kernels K = 1	K = 2	K = 3
4a	C = 1	.20	.22	.28
	C = 2	.21	.29	.70
	C = 3	.32	.59	2.07
4b	C = 1	.20	.22	.28
	C = 2	.25	.41	.89
	C = 3	.32	.71	1.95

[a](Mean Presentation Time − .5 Second) / Words.

per word and plotting log 10 Y against C^n, we obtain values of the fitted constant n as follows:

Experiment 4a: 5.3, 3.15, and 1.35 for K = 1, 2, 3.
Experiment 4b: 1.65, .75, and .30 for K = 1, 2, 3.

Values for the fitted constant g (g = n/K) are:

For Experiment 4a, 5.3, 1.57, and .45, for K = 1, 2, 3.
For Experiment 4b, 1.65, .37, and .10, for K = 1, 2, 3.

We noted earlier that when the measure is mean presentation time, g is a Type 3 function of K. When the measure is mean presentation time per word as this is given in Table 7, g is a Type 2 (simple exponential) function of K. While, in consequence, mean presentation time per word accelerates at a lower rate with increasing K than does mean presentation time, the basic function remains the same—$Y' = ae^{bX^n}$.

Passage Length in Words

While mean presentation time is not a linear function of number of words to be recalled (W), the number of words in a sentence tends not to deviate greatly from being a linear function of K x C. Therefore, the possibility exists that the presentation time data could be described by an equation having just one parameter, W.

DISCUSSION

The Immediate Data. The conditions of data collection in this study—a few stimuli per treatment value, a single subject—ordinarily work against the generation of coherent data. Thus, the data of Experiments 4a and 4b are meager, and the foregoing analyses may be more comprehensive than is warranted. However, taking the data at face value, the following propositions may be worth entertaining:

(1) Sentence recall is similar for sentences containing coordinate clauses and sentences containing non-coordinate clauses. The relative difficulty of sentences containing non-coordinate clauses may be somewhat greater, if only because such sentences contain more words.

(2) When the parameters C and K are used to measure sentence length, both total presentation time and presentation time per word will be monotonic, positively accelerated, Type 3 (parameter-exponented exponential), increasing functions of sentence length.[1]

Comparison of Experiments 2b and 4a. The clauses of Experiment 2b, although in list format and punctuated as distinct sentences, could be interpreted as belonging to a single sentence composed of a series of coordinate clauses. These clauses consisted of a single kernel sentence of form "N V N". The Experiment 2b data are most comparable with that part of the Experiment 4a data wherein K = 1. The two sets of data were described as follows:

Experiment 2b: $Y' = 6.0C^{2.05}$

Experiment 4a (K = 1): $Y' = 1.18e^{.25C^{1.48}}$

These equations differ primarily in the following sense: (a) For the Experiment 2b data, which are represented by a Type 1 (parabolic) equation, the acceleration of the function approaches linear acceleration. (b) For the Experiment 4a data, which are represented by a Type 3 equation, the function is extremely positively accelerated. When the measure is recall time per unit of material, the difference in acceleration remains. Thus, the rate of increasing difficulty of recall with increasing amount of material is considerably greater for the Experiment 4a materials than for those of Experiment 2b.

The possible reasons for the difference in functions include: (a) Clause form "N V N" in Experiment 2b and clause form "N is A" in 4a; (b) 16 lightly practiced and perhaps moderately motivated subjects in Experiment 2b and one highly practiced and well-motivated subject in 4a; (c) two-syllable content words in Experiment 2b and one-syllable content words in 4a.

[1]Recall relations may vary from one set of sentence types to another. It is possible that for one set of sentence types, recall will be a function of W, for another set a function of C and K, and for still another set a function of some other parameters.

It would seem that the paramount factors underlying the different functions would be amount of practice and degree of motivation. Data from a highly motivated, well-practiced subject should show the systematic form of relationships more than one who is not so well-motivated or practiced. For example, a relatively high constant latency incorporated into the Experiment 2b presentation time means could lead to accepting the function as Type 1 when in reality it was Type 3. These speculations, however, do not substitute for further study to ascertain definitively the form of relationship between time to learn and amount of material.

Earlier Research on Effects of Amount of Material. The question raised above and in Chapter 3 concerns the precise form (beyond monotonic, increasing with positive acceleration) of relationship that exists between presentation time per unit of material and the amount of material, when presentation intervals are set by the subject. The literature abounds with studies in which the findings suggest that, if the function is monotonic and increasing, at any rate it is negatively rather than positively accelerated. That is, the per unit cost in presentation time increases with negative acceleration as a function of amount of material. Negative acceleration means that the difficulty of per-unit recall decreases rather than increases as the amount of material increases.

Psychologists dealt most extensively with the question of the functional relation between amount of material and per unit presentation time during the period prior to World War I. This earlier work is summarized by Lyon (37) who himself contributed the most comprehensive findings of that period.

In a series of studies wherein he used himself as the sole subject, Lyon memorized materials every day before breakfast and after dinner throughout two 14-month periods. He studied the effects on cumulative presentation time to total written recall of amounts of four different kinds of material: (a) prose selections varying in length from 15 to 15,000 words, (b) poetry selections varying in length from one to 100 stanzas and from 25 to 2,500 words, (c) lists of nonsense syllables varying in length from eight to 300 items, and (d) number series varying in length from eight to 400 digits. He studied these effects under the condition of distributed practice—one reading daily—over the ranges of length cited above and under the condition of massed practice for more restricted ranges of length. In all of this work, duration of the presentation interval was under the subject's control.

There are many irregularities in the Lyon data. No doubt these are related to his basic design—to represent each treatment value of a study (e.g., 15,000 words of prose to be learned under the condition of distributed practice) with just one sample of material. Results were a function of the "representativeness" of single samples of material, and of the "representativeness" of single samples of Lyon's performance when recalling them.

For the four kinds of material employed by Lyon, given that practice is distributed, presentation time per unit of material tends to increase monotonically and with negative acceleration as a function of amount of material. This also tends to be true for prose, poetry, and digits when readings are massed; it tends not to be true for nonsense syllables under the condition of massed readings.

If the first treatment value for the nonsense syllable series is ignored, then the per unit presentation time under massed practice tends to increase exponentially and with positive acceleration with increasing number of nonsense syllables. Of the eight sets of data presented by Lyon, each dealing with an extensive series of amounts of material, seven support the view that per unit

presentation time increases with negative acceleration with increasing amount, and one supports the view that it increases with positive acceleration with increasing amount.

The Lyon data on prose were grouped to remove some of the variability, and mean presentation time per word was computed for the grouped data. The results appear in Table 8.

Table 8

Mean Presentation Time Per Word, Based on the Data of Lyon[a]

Average Words	Distributed Practice (sec)	Massed Practice (sec)
30	1.1	1.7
50	2.7	3.6
80	3.7	5.8
125	6.4	6.4
208	7.7	6.3
400	8.9	6.9
700	8.9	7.8
1033	9.1	9.5

[a]Reference 37.

When slight adjustments of the sort reflected in Table 7 are made, results in the present experiments show per unit presentation time in words to be increasing with positive acceleration with increasing amount of material. The same may be said for the Experiment 2a data as given, since each sentence of 2a contained five words.

The great discrepancy between the findings of Lyon and the present research could be due to the discrepancy in range of amounts of material used. Another possibility is that redundancy factors may have operated in Lyon's material to a much greater extent than in the materials used in the present research, which could be considered redundant only in their repeated use of function words and of grammatical form.

What factors underlie redundancy in prose are not yet clearly understood. It seems tenable, however, that when the amount of information per unit of material is at a maximum, then per unit presentation will be a function of amount of material such that the function's acceleration will be positive and at a maximal value. One way to attack the problem of identifying redundancy factors is to study the effects of potential factors of this sort upon the acceleration of functions of amount of material.

Thurstone (38) analyzed data obtained by Lyon, by Binet and Henri, by Ebbinghaus, and by Meumann. He found that, for a wide range of kinds of meaningless material, there is a tendency for number of repetitions of the material (readings, exposures, "trials") to increase approximately as the square root of amount of material—that is, with negative acceleration with increasing material. While the data presented in Table 8 do not quite agree with Thurstone's generalization, it may be that the generalization applies when such meaningful material is sufficiently well sampled to provide a normal level of redundancy. If this were the case, then a function on the order of $Y^2 = aX^n$ (where Y = number of repetitions or per unit presentation time and X = amount

of material) would be obtained, on the average, when spontaneously produced written material was used.

We may ask what happens to per unit presentation time when that most usual procedure of the psychology of verbal learning—fixed-interval presentation time—is employed. Hovland and Kurtz (39) found that when the individual items in lists of nonsense syllables were made familiar prior to list-learning, total trials to criterion increased approximately linearly with increasing number of nonsense syllables (over the range 6, 12, 24 syllables). Since each syllable was presented for two seconds, without pause between syllables of a list, the linear relation also holds for per syllable presentation time. Items of the six-syllable lists averaged 4.4 seconds; those of the 12-syllable lists, 22.8 seconds; those of the 24-syllable lists, 60.4 seconds of presentation time to criterion.

Chapter 6

SENTENCE RECALL AS A FUNCTION OF MODIFIER-HEAD ORDER IN THE NOUN PHRASE
(Study 5)

Abstract of Study 5

Two experiments were conducted on the effects upon sentence recall of placing a string of noun phrase modifiers before (Pre-H) or after (Post-H) the phrase head. In the first experiment, where modifiers of a string were *coordinate*, Pre-H and Post-H sentences proved to be about equally difficult to recall. In the second experiment, where words in the string were in *serial* modification, Pre-H sentences were significantly easier to recall than Post-H sentences. An analysis of presentation time per syllable suggests that sentence length alone may account for the differential recall of serial Pre-H and Post-H sentences.

BACKGROUND

Specifying certain grammatical characteristics that published prose will manifest is a perquisite of editorial office. On this authority, Bentley (40) and, more recently, Sanford (41) have discouraged the practice of using nouns attributively—as in "impression formation process." Presumably other editors have lobbied against this practice over the years. Such a position might be based on one or more of three considerations: It may be that the attributive use of nouns (a) offends personal taste, (b) violates a rule of valid grammars, or (c) degrades understanding relative to that which would occur if alternative means of expression were employed. Perhaps (a) could be set aside if (b) and (c) could be rejected.

With reference to (b), noun phrases whose heads are preceded by one or more attributive nouns are not considered ungrammatical in contemporary descriptive grammars. Thus, Francis (42, p. 297) says, "The reader must clear his mind of the notion that it is somehow more 'correct' for an adjective to modify a noun . . . than it is for a noun to modify a noun . . . ." With reference to (c), the relative understandability of contents stated in alternative grammatical forms can be treated empirically. In the experiments to be reported, understandability was inferred on the basis of sentence recall performance.

The attributive use of nouns is only one facet of the general problem of attribution in noun phrases. Consider the phrase "Soils Protection Agency." In the terminology of Nida (43), the first noun is a secondary post-determiner attributive, the second a primary post-determiner attributive, and the third a phrase head. Such a construction has the same syntactic status as a phrase containing, in order, an adverb, a verbal adjective, and a noun—for example, "recently published article." In both phrases a secondary attributive modifies a primary attributive, which modifies a phrase head. Because both exhibit serial modification (A modifies B, B modifies C), both have a left-branching tree-diagrammatic structure.

By contrast, a phrase wherein a series of primary multiple accumulative adjectives modify the head, as in "eminent, critical scientist," illustrates coordinate modification (A modifies C, B modifies C). Such a phrase has a right-branching structure.

Left-branching and right-branching phrase structures are illustrated in Figure 4. The right-branching structure asserts "Four clerks, smart clerks, young clerks, office clerks." The left-branching structure asserts "Panel Truck, Truck Driver's, Driver's Union, Union rules."

Left- and Right-Branching Phrase Structures

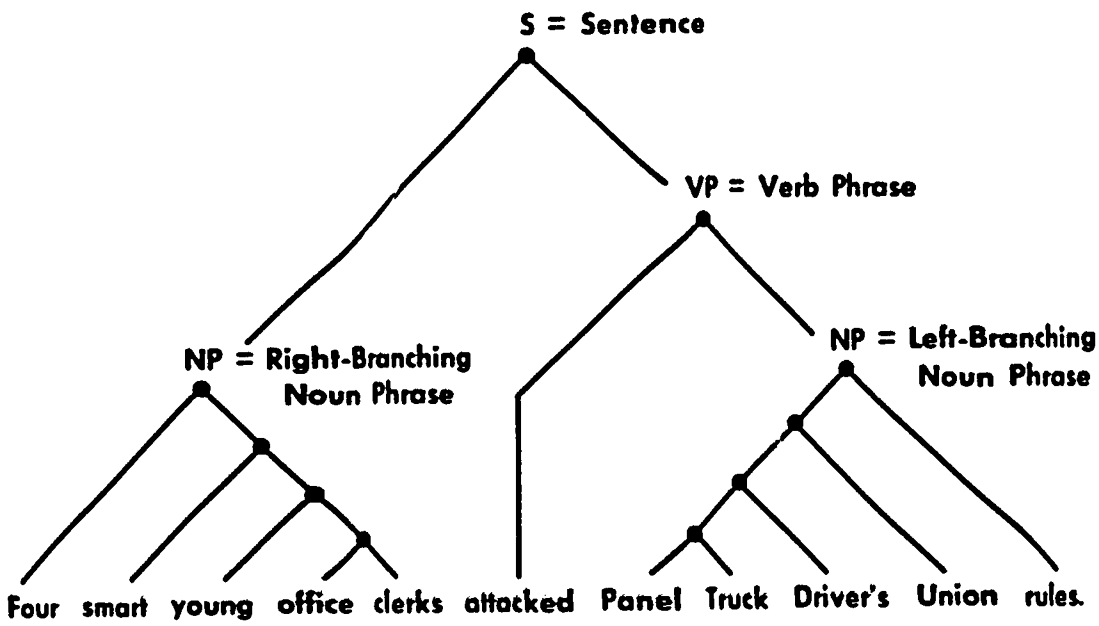


Figure 4

Some linguists (e.g., Gammon, 44; Yngve, 45) believe that left-branching constructions impose a greater burden on the hearer or reader than do right-branching. Others (e.g., Lees, 46; Miller and Chomsky, 25, p. 475) disagree. Members of the second group believe that the extent of branching—a function of number of attributives in the phrase—is more crucial to comprehension than its direction. According to this view, left-branching might in certain instances prove more efficient, as when the left-branching "Soils Protection Agency" is restated as right-branching "Agency for Protection of Soils," since the second structure contains more branches. Phrase length, whether in words or syllables, also increases when the left-branching phrase containing a string of serial modifiers is restated as a right-branching phrase.

When attributes are in serial modification, the phrase structure is left-branching if attributes precede the phrase head. However, when modifiers are coordinate, pre-head modification yields a right-branching structure. It is conceivable that modifier-head order effects will not be the same for coordinate and serial modification.

Concerning coordinate modification, Lambert and Paivio (47) reported that the French system of ordering nouns and adjectives—noun + adjectives—is superior to the English system—adjectives + noun—when English-speaking subjects learn lists containing blocks of words consisting of a noun and three primary accumulative adjectives, as in "eminent critical efficient scientist." Interestingly enough, their subjects made fewer errors to criterion when the noun came first, as in "scientist eminent critical efficient." Whether this finding is pertinent to the recall of full English sentences is not presently known.

In a follow-up study wherein subjects learned the response members of paired-associates lists whose items consisted of an adjective and a noun,

Paivio (48) found that the learning of response members was best when items of a list were in noun-adjective order. That is, when response members were adjectives and stimulus members nouns, subjects recalled more response members than when response members were nouns and stimulus members adjectives. This finding held whether the nouns employed were abstract or concrete.

The present research explores the effects upon the recall of English sentences of (a) coordinate modification in the noun phrase with modifiers preceding or following the head (Experiment 5a), and (b) serial modification in the noun phrase with modifiers preceding or following the head (Experiment 5b).

METHOD

Materials

Experiment 5a. The sentences of Experiment 5a contained nine content words—a verb and two noun phrases, each consisting of a noun and three primary multiple accumulative adjectives—together with such function words as were syntactically necessary. The content words of a noun phrase were an intact block from the materials used by Lambert and Paivio.

Each sentence was expressed in two forms: (a) AN (adjectives preceding the noun) and (b) NA (adjectives following the noun). Six sentences written in the alternative forms AN and NA were used. The following illustrate the alternative forms:

> Form AN: The eminent, critical, efficient scientist possessed a pronounced, harsh, artificial accent.
>
> Form NA: The scientist—who was eminent, critical, and efficient—possessed an accent which was pronounced, harsh, and artificial.

Experiment 5b. The sentences of Experiment 5b also were expressed in two forms: (a) MH (modifiers preceding the head) and (b) HM (modifiers following the head). A sentence in MH form consisted of a subject noun phrase, a verb, and two other noun phrases of varying syntactic function which may be termed the first noun phrase of the predicate and the second noun phrase of the predicate.

A sentence contained 12 content words—a verb and 11 other content words distributed over the noun phrases as follows: (a) subject noun phrase—secondary attributive (noun), primary attributive (noun), noun head; (b) first noun phrase in the predicate—secondary attributive (adverb of degree), two primary accumulative attributives (varying in finer grammatical classification), noun head; (c) second noun phrase in the predicate—secondary attributive (adverb of degree), two primary accumulative attributives (adjective plus noun), noun head. Six sentences written in the alternative forms MH and HM were used. The following illustrate the alternative forms:

> Form MH: The Operations Management Board revealed a relatively complex accounting system to a partially blind Sales Director.
>
> Form HM: The Board for Management of Operations revealed a system of accounting which was relatively complex to a Director of Sales who was partially blind.

Apparatus. The same apparatus was used as in Study 1. However, sentences appeared on 5 x 8 cards, rather than on slides.

Design. Each of eight subjects learned three of six sentences used in a study in one form (AN, MH) and the other three sentences in an alternative form (NA, HM). The design of each of the studies was a 2x2 Latin square (but see below) wherein one group of four subjects received one set of three sentences in form A_1 and a second set in form A_2, while a second group of four subjects received the first set of sentences in form A_2, and the second set in form A_1. The series of six sentences used in an experiment was preceded by a practice series consisting of a sentence in form A_1 and a sentence in form A_2. The sentences of each form were grouped together in the experimental series.

Two measures of proficiency were used in the analysis of the data: cumulative presentation time and cumulative response time over exposures to a criterion of one perfect immediate recall of the sentence. Criterion was reached when the subject made no syntactic errors and placed all content words in the sentence in their proper order. In assessing the subject's performance, function word synonyms were considered correct; content word synonyms were not.

In both experiments it seemed possible that any differences in difficulty of sentence forms might be attributable to differential difficulty of three-sentence sets or to differential ability of four-subject groups. We reasoned that if such differences existed between sentence sets or groups of subjects (Su), then, if the studies were replicated with conditions reversed (so that the first group of subjects now received the second set of sentences in form A_1 and the first set in form A_2 and the second group of subjects the reverse), the F ratio of treatments (Tr) by sessions (Se) interaction to Tr x Se x Su interaction in the analysis of variance would be significant. This possibility was tested in both studies. The reversed-conditions "replications" were conducted one week after the original data were collected.

Subjects. The eight subjects were volunteers from the local pool of HumRRO civilians and military research assistants. Subjects ranged in age from 22 to 34 years and in education from one to five years of college. The same subjects were used in both experiments.

Procedure. Sentences were presented on 5x8 cards. The rest of the procedure was the same as that of Study 1.

RESULTS

Treatment means, standard deviations, and F ratios for cumulative presentation time and response time to criterion, by experiment and session, are presented in Table 9. For the original (Session One) data, only the F ratios for differences in mean presentation time between sentences recalled in forms AN and NA failed to be significant at the .05 level; while none of the F ratios is significant for the replication (Session Two) data, all differences favor prehead modification (AN, MH).

Mean presentation time and response time measures, by session, are plotted in Figure 5 for Experiment 5a and in Figure 6 for Experiment 5b. Inspection shows that treatment pairs of "trends" for each measure in each experiment tend toward being parallel.

None of the F ratios of treatments by sessions interaction to treatment by sessions by subjects interaction (ms_{TrSe}/ms_{TrSeSu}) approached significance; for all but the response time pair of Experiment 5b, the treatments by sessions interaction F ratio was less than one. Hence, the possibility that differences

Table 9

Presentation and Response Times, by Session: Experiments 5a and 5b

Measure	Experiment 5a[a]				Experiment 5b[b]			
	Session 1		Session 2		Session 1		Session 2	
	AN	NA	AN	NA	MH	HM	MH	HM
Presentation Time (seconds)								
Mean	36.9	37.5	29.2	30.1	46.6	66.1	42.2	56.6
SD	14.2	12.3	12.8	10.3	25.0	25.5	21.1	34.1
F^c	<1		<1		24.8**		4.05	
Response Time (seconds)								
Mean	25.4	32.1	21.0	25.7	36.0	49.0	31.0	36.9
SD	13.2	14.5	13.9	12.6	13.7	18.4	14.9	13.4
F^c	6.72*		3.51		10.4**		1.52	

[a]AN, adjectives preceding the noun; NA, adjectives following the noun.
[b]MH, modifiers preceding the head; HM, modifiers following the head.
[c]$df = 1, 7$; F of $-5.59 = p$ .05 (*); F of $-12.25 = p$ .01 (**).

Session Effects for Mean Presentation and Mean Response Time: Experiment 5a

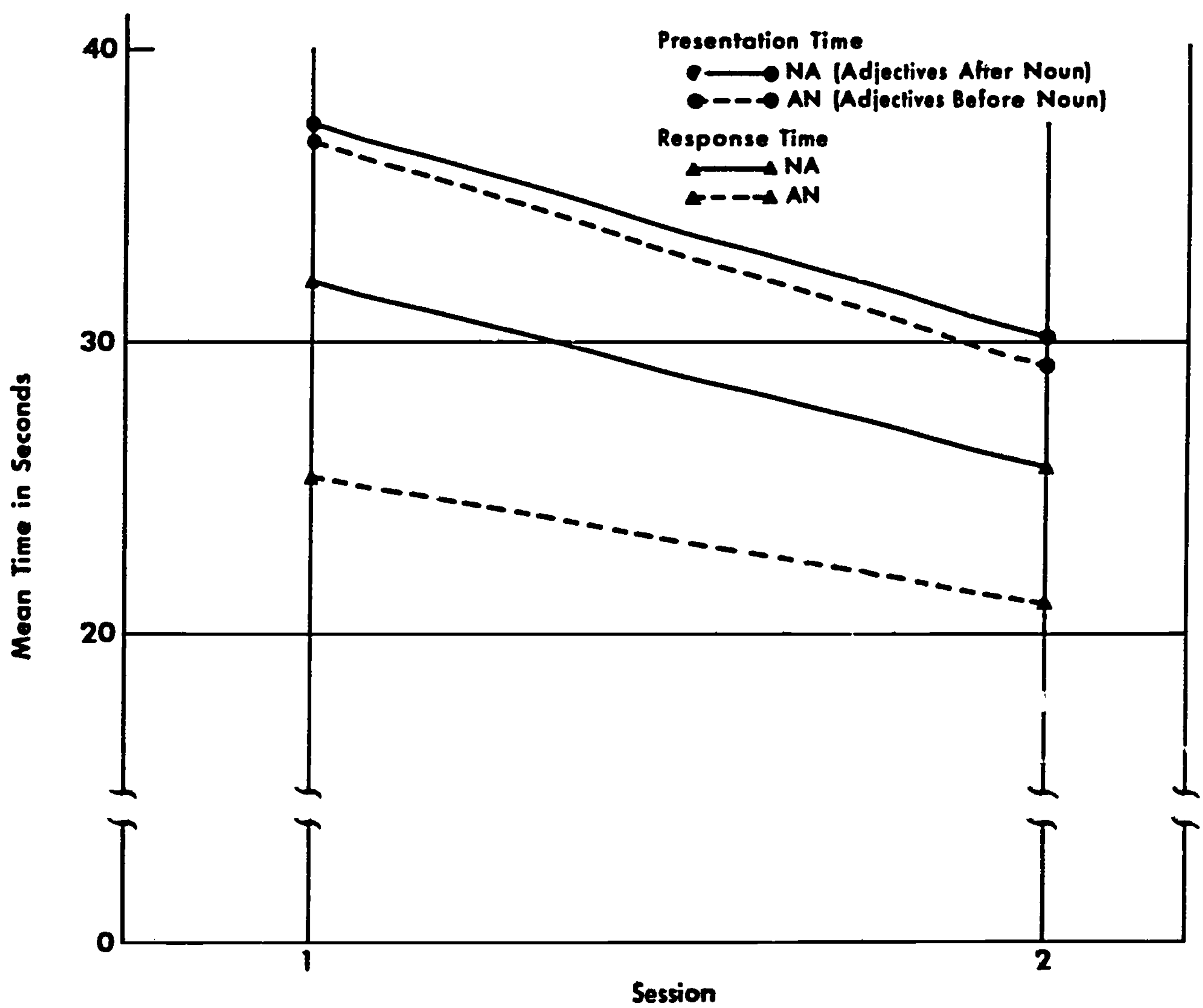


Figure 5

Session Effects for Mean Presentation and Mean Response Time: Experiment 5b

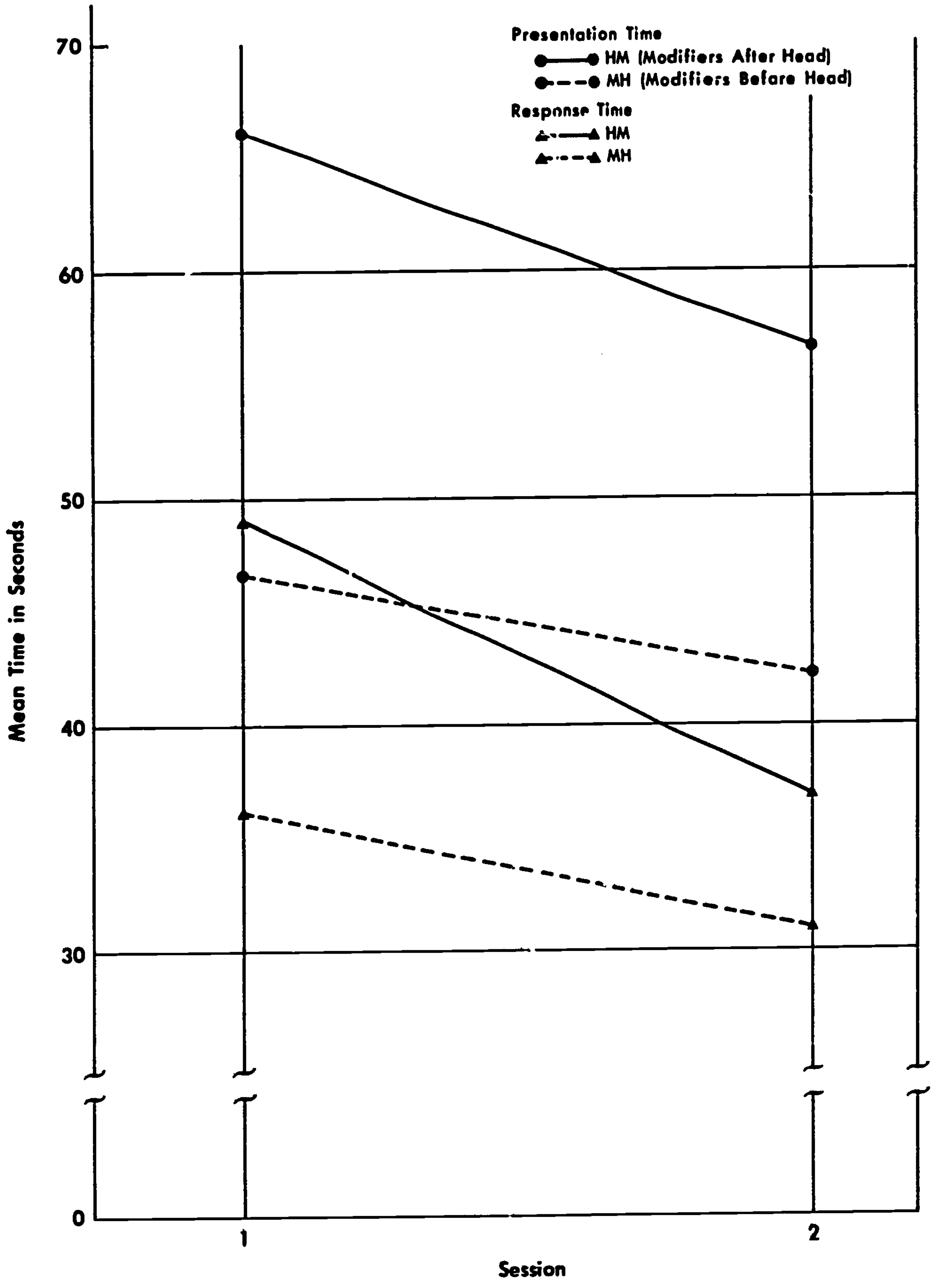


Figure 6

in difficulty of sentence forms might be attributable to differential difficulty of three-sentence sets or to differential ability of the four-subject groups is ruled out. Because none of the interaction $\underline{F}$ ratios was significant, the data across sessions were pooled. Treatment means, standard deviations, and $\underline{F}$ ratios for cumulative presentation time and response time to criterion, by study for the pooled data, are presented in Table 10.

Table 10

Presentation and Response Times, for Combined Sessions: Experiments 5a and 5b

Measure	Experiment 5a		Experiment 5b	
	AN	NA	MH	HM
Presentation Time (seconds)				
Mean	33.1	33.8	44.4	61.4
SD	11.7	10.7	22.5	26.8
F^a	<1		26.4**	
Response Time (seconds)				
Mean	23.2	28.9	33.5	42.9
SD	12.2	12.8	12.6	15.3
F^a	14.8**		7.86*	

[a] $df = 1, 7$; F of $5.59 = p$ .05 (*); F of $12.25 = p$ .01 (**).

We noted earlier that sentences in NA and HM form are longer than sentences in AN and MH form because additional function words are needed in the NA and HM forms. Any differences between AN and NA or MH and HM treatments might well be due entirely to the time it takes to read a sentence during presentation—silent reading time per "trial", or SR—and to the time it takes to report the sentence orally during responding—oral reading time, or OR. In order to test this possibility, the following procedure was devised and applied after data collection:

(1) Since the eight subjects fell into two distinct populations on the basis of performance in Experiments 5a and 5b, one subject from Group I ($\underline{N}=5$) and one from Group II ($\underline{N}=3$) were selected as typical of their respective groups.

(2) Oral reading rates for study materials were established for these two subjects and their rates generalized to other members of their groups.

(3) Woodworth (49) reports a study by Huey in 1908 which found that university students reading an interesting novel will average 3.6 words per second when reading at an ordinary rate orally and 5.6 words per second when reading at an ordinary rate silently. Hence, silent reading rates were established according to the relation SR (words) = .643 OR (words). This relation asserts that a given unit of material will require just over 1.5 times as much time to read orally as to read silently.

(4) Each subject's estimated silent reading rate times number of presentations to criterion was subtracted from his cumulative presentation time, and his estimated oral reading rate times number of presentations to criterion was subtracted from his cumulative response time. The reworked data—for Session 1 only—were based upon per "trial" rate values given in Table 11.

Table 11

Silent and Oral Reading Rate Values Per "Trial," Applied to Session 1 Data of Experiments 5a and 5b

(seconds)

Treatment	Group I		Group II	
	Silent Reading	Oral Reading	Silent Reading	Oral Reading
AN	4.6	7.2	3.3	5.1
NA	5.1	7.9	4.1	6.3
MH	5.3	8.3	4.2	6.6
HM	6.6	10.2	5.3	8.2

Session 1 means, standard deviations, and F ratios for presentation and response time measures recomputed to exclude estimated silent and oral reading time are presented in Table 12. While the magnitude of all F ratios diminished in consequence of excluding estimated silent and oral reading time from the original Session 1 data (see Table 9), differences between the Experiment 5b treatment means continue significant at the .05 level.

Table 12

Presentation Time Minus Silent Reading Time and Response Time Minus Oral Reading Time: Session 1 of Experiments 5a and 5b

Measure	Experiment 5a		Experiment 5b	
	AN	NA	MH	HM
Presentation Time (seconds)				
Mean	29.0	28.7	35.4	49.8
SD	12.1	11.6	23.0	24.7
F^a	1		17.9**	
Response Time (seconds)				
Mean	13.9	18.2	18.5	24.2
SD	9.8	12.1	9.3	12.6
F^a	4.16		6.02*	

[a] $df=1, 7$; F of $5.59=p$ .05 (*); F of $12.25=p$ .01 (**).

If differential reading time alone will not account for such differences as were obtained, then the differential length of sentences might. Sentences whose modifiers follow the noun phrase head tend to be longer than those whose modifiers precede the head. Differential length is due entirely to the number of function words used. AN sentences contained two function words; NA sentences, eight. MH sentences contained four function words; HM sentences, 12.

Values for mean presentation time and response time per syllable for the combined sessions data are presented in Table 13. Only the difference between presentation time per syllable means for Experiment 5a is significant ($F = 17.0$, $12.25 = p$ .01, $df = 1, 7$). According to this measure of length, sentences in NA form require significantly less presentation time per syllable than those in AN form.

Table 13

Presentation and Response Times Per Syllable: Combined Sessions, Experiments 5a and 5b

Measure	Experiment 5a		Experiment 5b	
	AN	NA	MH	HM
Mean Syllables per Sentence	23	29	33	41
Mean Presentation Time per Syllable (seconds)	1.44	1.17	1.35	1.50
Mean Response Time per Syllable (seconds)	1.01	1.00	1.02	1.05

An alternative measure of length is reading intervals, based on the rate information supplied in Table 11. Values for mean presentation time, in silent reading intervals, and mean response time, in oral reading intervals, for the combined sessions data are presented in Table 14. Only the difference between means for presentation time, in silent reading intervals, for Experiment 5a is significant ($\underline{F}=10.3$, $5.59=\underline{p}\ .05$, $\underline{df}=1, 7$). This measure of length yields findings which parallel those for length in syllables.

Table 14

Presentation Time, in Silent Reading Intervals, and Response Time, in Oral Reading Intervals: Combined Sessions, Experiments 5a and 5b

(seconds)

Measure	Experiment 5a		Experiment 5b	
	AN	NA	MH	HM
Mean Presentation Time, in Silent Reading Intervals	7.8	7.0	8.8	9.8
Mean Response Time, in Oral Reading Intervals	3.6	3.9	4.5	4.5

On the basis of the data presented in Tables 13 and 14 and analyses presented in the text, it appears that the factor of length alone accounts for the superiority of MH sentence recall as this is reflected in Table 10. When length is considered, the superiority of AN sentence recall, as measured by response time, also disappears.

DISCUSSION

Experiment 5a. Results suggest that when the pattern of modification in noun phrases is coordinate, constructions involving pre-head modification and those involving post-head modification will be about equally easy to memorize.

Experiment 5b. Findings lend support to the view in linguistics that the direction of branching has less to do with efficiency in processing sentences than the degree of branching. When the task is to memorize sentences whose attributives form a pattern of serial modification, constructions involving pre-head modification will be easier to deal with than those involving post-head modification. Effects of pre-head attributive use of nouns were not studied independently of effects of other sorts of pre-head serial modification. While the data, therefore, cannot be used to refute the bias of an editor against using nouns attributively, neither do they lend comfort to such a view.

Length as a Predictor. These experiments and some earlier work we have done dealing with packaging variables tend to support the view that, given two grammatically equivalent ways of expressing a content, the shorter of these alternatives will be favored when the task is to memorize a sentence. Length may be defined in a number of ways. Mehler (50) used "number of transformations of an underlying kernel sentence" inherent in a one-clause sentence as a measure of length. Mehler's results are as clearly a function of a less elegant measure of length—number of syllables in the sentence—as they are of transformational complexity. On the basis of a study somewhat similar to Mehler's, Gough (51) reports that speed of verification of heard sentences varied with transformational complexity of the sentence, but that results could be attributed solely to differential length.

However length may be expressed, the results thus far obtained suggest that when a single sentence is to be memorized, effects of the length factor will outweigh those of grammatical factors per se. It remains to be demonstrated that this proposition holds for a wide variety of alternative, grammatically equivalent expressions, for units of content larger than the sentence, or for measures of proficiency other than memory measures.

Chapter 7

RECALL OF WORD LISTS AS A JOINT FUNCTION OF WORD LENGTH AND WORD FREQUENCY
(Study 6)

Abstract of Study 6

Five experiments on the joint effects of word length and word frequency on the recall of 15-word lists are reported. Four experiments employed three levels of word length (1, 2, or 3 syllables) and three levels of word frequency (5, 50, or 500 occurrences per 4.5 million words). The fifth employed the same three levels of word length and five levels of word frequency (5, 15, 45, 135, or 405 occurrences). Three of the 3x3 studies used one set of nine lists (Set A); the other 3x3 study, another set of nine lists (Set B). The 3x5 study used still another set of lists, 15 in number (Set C).

Results regarding word length effects tend to be consistent in the sense that recall time is a monotonic, increasing function of word length. The data do not appear to warrant a formal description of either the simple effects of word frequency or the joint effects of word length and word frequency. The results of these studies posed a major design problem of the sampling stimuli, since each was an attempt to evaluate the combined effects of word length and word frequency but each used only a single list.

BACKGROUND

Findings of Experiments 3a and 3b indicated that word length has a considerable effect upon recall. The list-learning literature contains some frequently cited studies—for example, Bousfield and Cohen (52), Hall (32), Sumby (33)—demonstrating that lists of high-frequency words are more easily recalled than lists of low-frequency words. Conversely, this literature contains some almost-never mentioned studies—for example, Gorman (53), Underwood and Postman (54), Waugh (55), Winnick and Kressel (56)—in which increasing frequency either had no effect or had an adverse effect upon recall.

The five experiments to be reported may be compared and contrasted on the basis of three factors: (a) the set of lists used (Sets A, B, C); (b) length-by-frequency levels (3 x 3, 3 x 5); (c) locus of presentation time control (subject, experimenter). The headings used in this chapter describe the experiments according to these three factors.

EXPERIMENT 6a: Set A Lists, 3x3 Levels, S-Controlled Time

It is known that length and frequency interact in their effects upon word recognition threshold. It seemed plausible that a similar interaction would hold for word recall. In Experiment 6a, we explored the joint effects of frequency and length upon recall, expecting to find a relationship in which an increase in frequency decreases recall time to criterion more for long words than for short words, while an increase in word length increases recall time more strikingly for low-frequency than for high-frequency words.

Nine lists of 15 words each were constructed, reflecting all combinations of three levels of word length—1, 2, 3 syllables—and three levels of word frequency. Word frequency was defined in terms of the Lorge Magazine Count (Thorndike and Lorge, 18). The levels of word frequency employed were 4-6, 40-60, and 400-600 occurrences per 4.5 million words of running text. (Word frequency levels will hereafter be referred to as the mean values for these intervals—5, 50, and 500 occurrences.) List items were randomly selected from the Magazine Count except that certain classes of words such as proper names and abbreviations were not used.

Nine HumRRO civilian personnel drawn from the same pool as the Study 5 subjects were used in Experiment 6a. Lists were presented by the whole method of presentation until a criterion of one perfect recitation was met—that is, all words recalled, in any order. Presentation time and response time were under subject's control.

Analyses were performed for four measures: (a) cumulative presentation time to criterion, (b) cumulative response time to criterion, (c) total time—presentation time plus response time to criterion, and (d) "trials" to criterion.

For all time measures, word length main effects were significant at the .05 level, while frequency effects were not. As expected, time to criterion increased with increasing word length. For both response time and total time measures, the interaction between frequency and word length was significant at the .05 level; the word length effect was greatest for the high-frequency lists. Neither word length or frequency was a significant source of variance when the measure was "trials" to criterion.

Thus, the interaction obtained was the opposite of what had seemed probable, and no frequency effect was detected. Because a replication seemed in order, Experiment 6b was conducted.

EXPERIMENT 6b: Set B Lists, 3x3 Levels, S-Controlled Time

Subjects and procedure were the same as in Experiment 6a. Nine new word lists were constructed representing the same combinations of word length and frequency as those used in Experiment 6a.

Although, as in Experiment 6a, time to criterion tended to increase as word length increased, word length main effects did not reach the level of significance for any of the time measures or for "trials" to criterion. Word frequency main effects were significant at the .01 level for presentation and total time measures. Presentation time and total time to criterion decreased as frequency increased. There were no significant interaction effects.

An interpretation of the contradictory results of Experiments 6a and 6b is somewhat hindered by the fact that subjects came to the second of these experiments better practiced in list-learning than they were at the outset of the first. Virtually all lists of Experiment 6b were learned more rapidly than their counterparts in 6a. However, the form of the effect was puzzling. The increase in efficiency of recall was far more pronounced for lists of high-frequency words than for those of low frequency. If improved efficiency can be attributed only to practice effects, then the effects of practice increased as frequency increased.

The possibility exists that the contradictory results may have arisen from faulty sampling of lists used to represent the treatment combinations. In both experiments, each treatment combination was reflected in a single list. It is apparent that we were not, at the conclusion of Experiment 6b, sufficiently

punished by unruly findings to treat sampling from word frequency lists more thoroughly. Hence, we continued to use the one-list methodology.

EXPERIMENT 6c: Set A Lists, 3x3 Levels, S-Controlled Time

It seemed possible that earlier results were tied to the vocabulary sophistication of subjects used in these studies. Hence, Experiment 6a was replicated using naive subjects whose vocabulary skills presumably were lower. Subjects were nine first-year Army enlisted men, none of whom scored below 100 on the VE (Verbal) Subtest of the Army Classification Battery. The same materials and procedure were used as in Experiment 6a.

For cumulative presentation time to criterion, the word length main effect was significant at the .05 level. Presentation time to criterion increased with increasing word length. No other significant main effects or interactions were obtained. Pronunciation errors and comments made by the subjects left no doubt that the words in the low-frequency lists were far less familiar to them than to the subjects in Experiment 6a. Evidently vocabulary knowledge was not at the root of our failure to find frequency effects in Experiment 6a, since such effects did not materialize in 6c.

EXPERIMENT 6d: Set A Lists, 3x3 Levels, E-Controlled Time

Experiment 6d paralleled 6c except that presentation time was fixed at 30 seconds per trial. A different sample of subjects was used, drawn from the same population as in Experiment 6c. Experiment 6d, then, is also a replication of 6a although, like 6c, using subjects with a lower degree of vocabulary sophistication and, unlike 6c, using fixed-interval presentation time.

Results were more complicated than those obtained for Experiment 6c. Word length main effects were significant at the .005 level for all measures. Trials, response time, and total time to criterion increased as word length increased.

Word frequency main effects were significant at the .01 level for trials to criterion and at the .05 level for total time to criterion. Trials and total time to criterion decreased slightly as mean list frequency increased from 5 to 50 and increased sharply as mean list frequency increased from 50 to 500. Thus, the lists of highly familiar words were the most difficult to recall.

A significant interaction ($p < .005$) between length and frequency was found for trials to criterion. Increasing word length increased trials to criterion more for high-frequency than for low-frequency words; however, a dramatic increase in trials to criterion as mean list frequency increased from 50 to 500 occurred only for the two- and three-syllable lists. Analyses of response oscillation rates and of frequency of extra-list intrusions in recall suggested a high level of response interference in the two-syllable and three-syllable lists at the 500 mean frequency level.

SUMMARY OF EXPERIMENTS 6a THROUGH 6d

Results, at the end of four experiments, were a mixture that seemed not to fit together. Only word length effects across studies appeared coherent; the

occasional failures of length effects to reach significance were not too disturbing in view of the small number of subjects used in each study and the consistency of trends. Other results were disappointing.

The data were examined further in an effort to understand the discrepancies among findings of these four studies and those of other investigators regarding the relationship between frequency and the recall of word lists. The following pertinent discoveries were made:

(1) In all four studies, mean response rate increased with increasing word frequency. Clearly, in a study such as Sumby's (33), where the subject is allowed a very short fixed interval in which to respond, superior recall of high-frequency items will be favored because of their shorter latency.

(2) In terms of items recalled per unit of presentation time, first-trial retention was superior for high-frequency lists in all four studies. This initial advantage was maintained to criterion in Experiment 6b, wherein item presentation rate was low and subjects verbally fluent and highly practiced. In Experiment 6d, under circumstances likely to promote interference—that is, high item presentation rate, lower verbal fluency and less practice among subjects—the first-trial frequency effect was rapidly reversed on subsequent trials. Since high-frequency words evoke stronger associative responses than low-frequency words, it is not surprising that they should be more susceptible to interference effects in recall.

EXPERIMENT 6e: Set C Lists, 3x5 Levels, E-Controlled Time

It appeared that, before any conclusions were drawn, one more study should be done using a larger number of subjects and more care in constructing lists (although, unfortunately, we still did not see the necessity of increasing the number of lists associated with each treatment combination). It appeared tenable that purely formal similarities between items within lists might have contributed to interference effects in earlier studies.

Method

Materials. Fifteen lists of 15 words each were constructed at three levels of word length (1, 2, or 3 syllables) and five levels of mean word frequency (5, 15, 45, 135, or 405 occurrences per 4.5 million words of running text in the Magazine Count). No two items within a list began with the same digram or ended in the same trigram. No syllable was repeated in a given list. Mean phonemic word length and mean number of letters per word were controlled across frequency levels. No function words, abbreviations, proper names, or hyphenated words were used.

Procedure. Procedure followed that used in Experiment 6d. Each subject learned all 15 lists plus a practice list over a period of two days.

Subjects. Subjects were 30 first-year Army enlisted men with scores at or above 100 on the VE (Verbal) Subtest of the Army Classification Battery.

Results

Mean Response Rate. Mean response rate is defined as the number of responses (including incorrect responses, such as intrusions) to criterion divided by cumulative response time to criterion. Both word length and word

frequency were significant sources of variance in mean response rate. For word length, $F = 47.7$, $df = 2, 58$, $p < .005$; for word frequency, $F = 26.6$, $df = 4, 116$, $p < .005$. There was no interaction between length and frequency. Response rate was a linear increasing function of the logarithm of word frequency and a decreasing function (probably linear) of word length.

Response Time to Criterion. The word length main effect was significant ($F = 24.3$ $df = 2, 58$, $p < .005$); so was the frequency main effect ($F = 13.0$, $df = 4, 116$, $p < .005$). No significant interaction was found. Response time to criterion decreased as frequency increased and as length decreased.

Trials to Criterion. The length main effect ($F = 19.1$, $df = 2, 58$), frequency main effect ($F = 8.77$, $df = 4, 116$), and interaction between frequency and word length ($F = 3.09$, $df = 8, 232$) were significant at the .005 level. More trials were required to reach criterion on three- than on one- or two-syllable lists. Lists at mean frequency levels of 135 and 405 required fewer trials than those at levels 5, 15, and 45. Differences between the two high-frequency levels and among the three low-frequency levels were negligible. The interaction effect apparently is due to the fact that frequency had no significant effect upon the recall of lists of two-syllable words. In short, the interaction defies interpretation.

Mean Trial of First Recall. This measure may require explanation. Underwood and Postman (54) have proposed that the trial on which a response is first given may be used to measure the speed with which the first stage of learning—response integration—takes place. They suggest that the mean trial of first recall of items of a list presented according to fixed-interval procedure should reflect the frequency of items with greater sensitivity than a trials to criterion measure.

An analysis of variance of mean trial of first recall yielded significant main effects for word length ($F = 41.7$, $df = 2, 58$, $p < .005$) and word frequency ($F = 11.5$, $df = 4, 116$, $p < .005$), and a significant interaction between length and frequency ($F = 2.38$, $df = 8, 232$, $p < .05$). Mean trial of first recall decreased as frequency increased and as length decreased. The interaction is the type we had originally expected to find when this series of studies was initiated. The length effect was most pronounced for low-frequency words, and the frequency effect most pronounced for the longer words.

Discussion

In general, main effects and interaction F ratios for Experiment 6e conform to what we would expect. Trends based upon plots of mean response rate (Y) against word length (X) and word frequency (Z) also conform to expectations. Trends based upon plots of trials to criterion (trials x 30 seconds= cumulative presentation time to criterion) do not conform to expectations.

Preliminary Methodological Evaluation of Sampling Problem Posed by Single Lists

Treatment means for the experiments involving subject-controlled presentation time—6a, 6b, and 6c—are presented in Table 15, and those for the fixed-interval experiments—6d and 6e—in Table 16. Entries are for cumulative mean presentation time to criterion; Table 16 entries were obtained by multiplying trials to criterion by 30 seconds, the length of each fixed interval.

Experiments 6a, 6c, and 6d used Set A lists as stimuli, Experiment 6b used Set B lists, and Experiment 6e used Set C lists. If those Experiment 6e lists

Table 15

Presentation Times: Experiments 6a, 6b, and 6c (Variable Interval)
(seconds)

Number of Syllables	Experiment 6a Word Frequency				Experiment 6b Word Frequency				Experiment 6c Word Frequency			
	5	50	500	Mean	5	50	500	Mean	5	50	500	Mean
1	234	216	152	201	215	164	158	179	558	670	456	561
2	219	218	246	228	251	182	165	199	696	627	624	649
3	292	301	252	281	230	195	179	201	700	706	731	712
Mean	248	245	216		232	180	167		651	667	604	

Table 16

Presentation Times: Experiments 6d and 6e (Fixed Interval)
(seconds)

Number of Syllables	Experiment 6d Word Frequency				Experiment 6e Word Frequency					
	5	50	500	Mean	5	15	45	135	405	Mean
1	273	280	267	273	334	311	360	291	233	306
2	303	267	447	339	336	309	294	281	298	304
3	357	343	540	413	400	441	429	318	360	390
Mean	311	297	418		357	354	361	297	297	

wherein word frequency equals 5, 45, and 405 occurrences per 4.5 million words are considered as equivalent to Set A and Set B lists whose words represent frequencies of 5, 50, and 500 occurrences, then Tables 15 and 16 supply data for three lists associated with each of the nine combinations of three word lengths (1, 2, and 3 syllables) and three frequencies (5, 50, and 500 occurrences).

Procedure and the population of subjects from which samples were drawn varied between experiments, with effects that could only increase superfluous variance when results across studies are pooled. Nevertheless, let us see what happens when means are based on three lists, rather than one, per combination of word lengths and frequencies.

Since Set A lists were used in connection with Experiments 6a, 6c, and 6d and we have no a priori basis for selecting the results of one of these studies as typical for the Set A lists, presentation time means for each of these studies were pooled with those of the other two studies—Experiments 6b, using Set B lists, and 6e, using those Set C lists reflecting frequencies of 5, 45, and 405 occurrences. The results of pooling are presented in Table 17. Entries are means of the means for three studies.

Column and row means reflect main effects and entries to the left of and above these mean values associated with 3 x 3 surfaces. It is evident that column and row means for the pooled 6a-6b-6e and for the pooled 6c-6b-6e data at least order properly in terms of what would be expected. Those for the pooled 6d-6b-6e data do with one exception, the mean for entries of the "405 or 500" frequency column.

In general, 6a-6b-6e entries are ordered in a consistent manner, whether across syllables or frequency. The departures from expectation are not great.

Table 17

Pooled Means for Presentation Times: Study 6 Experiments[a]
(seconds)

Number of Syllables	6a-6b-6e Data Word Frequency				6c-6b-6e Data Word Frequency				6d-6b-6e Data Word Frequency			
	5	45 or 50	405 or 500	Mean	5	45 or 50	405 or 500	Mean	5	45 or 50	405 or 500	Mean
1	261	247	181	230	369	398	282	350	274	268	219	254
2	269	231	236	245	428	368	362	386	297	248	303	283
3	307	308	264	293	443	443	423	436	329	322	360	337
Mean	279	262	227		413	403	356		300	279	294	

[a]Each of the three sets of pooled data is based upon one experiment using Set A lists (6a, 6c, or 6d), one using Set B lists (6b), and one using Set C lists (6e).

Entries for 6c-6b-6e tend toward ordering consistency with the marked exception of the entry for the "45 or 50" frequency, one-syllable condition. At least three entries for the 6d-6b-6e pooled data are of magnitudes inconsistent with expectation.

Since Experiment 6d effects were quite different from those of the other experiments using Set A lists (Experiments 6a and 6c), we cannot ascribe the entire problem of contradictory findings reflected in the series of experiments to inadequate stimulus sample size. However uninterpretable the findings of Experiments 6a, 6c, and 6d might have been, there should have been a certain consistency in the ordering of means if the problem rested entirely with inadequate stimulus sample size (one list per treatment combination).

However, it is clear from the results presented in Table 17 that there is a good deal to be said for the notion that inadequate sample size was a major barrier to obtaining coherent data in this series of studies. We may speculate, on the basis of the pooled results reflected in Table 17, that, for reasons unknown, the Experiment 6d data do not belong to the population of data that these studies reflect, since the picture which the pooled 6d-6b-6e data poses deviates rather markedly from that posed by the two other sets of pooled data.

The comments in this section are put forth in support of a methodological alternative to that characterizing this series of studies. Little can be gleaned from the data gathered in these experiments. Determination of the nature of the functional relation holding between a measure of recall and word length and frequency, taken together, must await a study predicated on the use of several lists in association with each combination of levels of the two treatments.

It is possible that, if word frequency counts compiled prior to 1944 were indices of word familiarity at that time, cultural changes over the years may have rendered them less valid tools for gauging word familiarity in 1965. Our willingness to entertain that proposition to account for disordered findings is, however, quite low.

Chapter 8

IMPLICATIONS FOR FURTHER RESEARCH

The most obvious conclusion which can be drawn from the studies reported is that, of the four types of stimulus characteristics studied—load, length, packaging, and frequency—length has by far the clearest and most potent effects upon rote memorization.

LOAD

Study 1 dealt with the effects on paragraph recall of a measure of load—the ratio between total content word tokens and total word tokens. This particular measure of load (Content Word Ratio) was found to be linearly related to two different measures of length—number of syllables in the paragraph and estimated number of kernel sentences in the paragraph. On the basis of the relation between load and length holding in this particular study, it is evident that the measure of load used—CWR—will not predict any characteristics of paragraph complexity which cannot be predicted just as readily by a measure of length.

Study 1 results do not rule out the possibility that load factors will be found that account for much of the variance associated with processing written material. Although study of load factor effects would not have a high priority, further work involving load factors might be useful from time to time as the cumulative results of research point to a need for such studies.

PACKAGING

Study 5 was concerned with the effects upon sentence recall of a packaging factor—modifier-head order in the noun phrase. In such experiments sentence grammar—syntax—is manipulated while semantic meaning is held constant. In this and similar studies that use a measure of memorization as the dependent variable, the advantage that one of a pair of syntactic alternatives may hold over the other can be accounted for by considering the differential length of sentences employing the alternative forms. There seems as yet no clear evidence from our studies of the existence of alternative syntactic constructions wherein the factor of syntax itself uniquely contributes to differential difficulty of processing material in the alternative forms.

We have thus far dealt with only a few of the grammatically equivalent ways of packaging small units of content. It is too early to dismiss the possibility that syntax per se might be a major source of variance under certain conditions. It is possible that effects of syntactic factors will be clear only when larger units of content are employed, or that such effects would be detected when a measure of understanding is used but not when a measure of memorization is used.

We continue to believe that syntactic and intersentential grammatical factors that are amenable to manipulation when writing technical material will

eventually be found to contribute appreciably to complexity of the material and hence to the speed with which the material is understood. Further research involving packaging variables appears to offer considerable promise.

FREQUENCY

Both the results of Study 6 and other studies in the literature suggest that word frequency accounts for a certain amount of variance in processing words. It is apparent in Study 6, whatever its shortcomings, that the effects of frequency over a restricted range are not as great as those of word length. Perhaps these effects will be found to be a good deal more pronounced when the frequency range is increased to include the very low frequencies associated with more esoteric words.

Our interest in word frequency in print is based on the factor's utility as a predictor of familiarity. A number of alternative procedures can be used to estimate how familiar the subject will find a word. Unfortunately, the most sensible and straightforward approach—vocabulary testing geared to the specific vocabulary of interest—requires the development of appropriate tests and time-consuming application of such tests.

Just what further research would be desirable with regard to effects of predictors of familiarity upon memorization and understanding is not yet clear. It appears that at least some additional study is needed to establish the functional relation between a predictor of familiarity and a measure of proficiency.

LENGTH

Studies 2, 3, and 4 show clearly that a wide variety of measures of length can be used to predict effects upon memorization of written material. This research suggests that the most promising single approach that could be made at this time to the problem of quantifying the complexity of written material would involve measures of length.

On the basis of findings to date, it would appear that, for nonredundant materials, effects of length upon memorization can be expressed by monotonic, positively accelerated increasing functions characterized by a high degree of acceleration.

Extensive additional research is needed on the quantification of effects of length factors upon memorization of written material.

REDUNDANCY

The factors associated with redundancy of written material are not self-evident, although we can speculate that such characteristics as word repetition, phrase repetition, and the like contribute to redundancy.

The Lyon data and the Thurstone generalization—touched on in Chapter 5—indicate that, when reasonably large amounts of ordinary written material are to be memorized, effects of length upon per unit recall will be monotonic and negatively accelerated. Our findings suggest that the function will be monotonic, positively accelerated, and increasing when redundancy—intuitively assessed—is low. Assuming this, we believe that functions of length that are negatively accelerated and increasing must be due to the material becoming increasingly redundant with increasing length. If redundancy were simply a matter of word

repetition, we could see how this might happen on the basis that a finite vocabulary can be used to write a passage of almost any length, including an infinitely long passage.

If monotonic, positively accelerated, increasing functions of length describe nonredundant material, then departures from these functions might be taken as proof of the operation of redundancy factors. Given such a test, it should be possible to reflect a number of potential redundancy factors in experimental materials and, on the basis of their effects upon memorization, to determine what factors need be considered in order to account for shifts in function when changing from nonredundant to redundant material.

A fruitful direction for research lies in a program wherein the difficulty of material will be a function of certain length factors whose effects will be modified in certain ways by certain redundancy factors operating upon the length factors. Eventually, it might be possible to establish empirically the pertinent factors underlying redundancy in written material and to quantify the effects of these factors upon memorization.

MEASURES OF PROFICIENCY

Questions arise from time to time concerning the adequacy of our proficiency measures as predictors of comprehension (understanding). We have been using a measure of memorization (M) wherein a response in the form of unassisted total rote recall is required at the end of an instructional session. The response of ultimate interest is a measure of comprehension (C). Intuitively, we all understand that M is primarily a matter of applying storage rules to a content, whereas C is a matter of applying storage rules plus certain other rules so that an initial content leaves storage in a form that is, linguistically, a paraphrasing or transformation of the content. Without suggesting a sophisticated analysis, we may say that M results from the application of Rules A, whereas C results from the application of Rules A plus Rules B.

We may ask whether $M = f(X)$ and $C = f(X)$ are sufficiently similar in form that, given $M = f(X)$, we could predict $C = f(X)$. Put another way, what function of M is C? Studies dealing with this important question would deserve priority in future research planning.

SUMMARY OF RESEARCH IMPLICATIONS

Research into length effects would appear to merit continued heavy emphasis. It also appears that research into the effects of redundancy factors eventually should be conducted on a scale commensurate with that contemplated for length factors.

Research into the interrelation of alternative measures of proficiency is needed to determine whether a straightforward measure such as memorization can be used to predict the complicated process of comprehension. The extent to which this question should be pursued would depend on initial findings.

A certain amount of further work concerning the effects of packaging variables also appears to be needed. At least one more study of frequency effects seems indicated. Studies into the effects of load factors should be undertaken to the extent that cumulative results of the research program point to their need.

LITERATURE CITED

LITERATURE CITED

1. Chall, Jeanne S. *Readability: An Appraisal of Research and Application,* Monograph 34, Bureau of Educational Research, Ohio State University Press, Columbus, 1958.

2. Gray, William S. "Reading," in *Encyclopedia of Educational Research* (3d edition), C.W. Harris (ed.), Macmillan Company, New York, 1960, pp. 1086-1135.

3. Hackman, Ray C., and Kershner, Alan M. *The Determination of Criteria of Readability,* Technical Report NR 153-024, Office of Naval Research, Washington, 1951.

4. University of Illinois, Division of Communication. *What Makes for Readable Writing and Reading Success,* Research Memorandum 10, Human Resources Research Institute, Maxwell AFB, Ala., July 1953.

5. Gray, William S., and Leary, Bernice E. *What Makes a Book Readable,* University of Chicago Press, Chicago, 1935.

6. Flesch, Rudolf. *How to Test Readability,* Harper & Brothers, New York, 1951.

7. Stevens, S.S., and Stone, Geraldine. "Psychological Writing, Easy and Hard," *Amer. Psychologist,* vol. 2, no. 7, July 1947, pp. 230-235, (Discussion by Saul Rosenzweig, pp. 523-525.)

8. Coleman, Edmund B. "Improving Comprehensibility by Shortening Sentences," *J. Appl. Psychol.,* vol. 46, no. 2, April 1962, pp. 131-134.

9. Coleman, Edmund B. "The Comprehensibility of Several Grammatical Transformations," *J. Appl. Psychol.,* vol. 48, no. 3, June 1964, pp. 186-190.

10. McCarthy, Dorothea. "Language Development in Children," in *Manual of Child Psychology* (2d edition), Leonard Carmichael (ed.), John Wiley & Sons, Inc., New York, 1954.

11. Harrell, L.E., Jr. "A Comparison of the Development of Oral and Written Language in School-Age Children," in *Monographs of the Society for Research on Child Development,* vol. 22, no. 3, Purdue University, Lafayette, Ind., 1957.

12. Hunt, K.W. *Differences in Grammatical Structures Written at Three Grade Levels, the Structures to be Analyzed by Transformational Methods,* Cooperative Research Project 1998, Florida State University, Tallahassee, 1964.

13. Baldwin, Alfred L., and Baum, Esther. "The Interruptability of Words in the Speech of Nursery School Children," in *A Basic Research Program in Reading,* Harry Levin, Eleanor Gibson, Alfred L. Baldwin, James J. Gibson, Charles F. Hockett, Henry N. Ricciuti, and George J. Suci (eds.), Cooperative Research Project 639, Cornell University, Ithaca, N.Y., 1963.

14. Suci, George J. "The Definition of Units in Language," in *A Basic Research Program in Reading,* Harry Levin, *et al.* (See Reference 13.)

15. Boder, David P. "The Adjective-Verb Quotient; A Contribution to the Psychology of Language," *Psychol. Rec.,* vol. 3, 1940, pp. 310-343.

16. Johnson, Wendell. "Language and Speech Hygiene: An Application of General Semantics," *Gen. Semantics Monogr.,* no. 1, Institute of General Semantics, Chicago, 1939.

17. Miller, George A. *Language and Communication,* McGraw-Hill Book Co. Inc., New York, 1951.

18. Thorndike, Edward L., and Lorge, Irving. *The Teacher's Word Book of 30,000 Words*, Columbia University Press, New York, 1944.

19. Fries, Charles C. *The Structure of English*, Harcourt, Brace & World, Inc., New York, 1952.

20. Miller, G.A., Newman, E.B., and Friedman, E.A. "Length-Frequency Statistics for Written English," *Information and Control*, vol. 1, 1958, pp. 370-389.

21. Taylor, Wilson L. "Application of 'Cloze' and Entropy Measures to the Study of Contextual Constraint in Samples of Continuous Prose," unpublished doctoral dissertation, University of Illinois, Urbana, 1954.

22. Aborn, Murray, Rubenstein, Herbert, and Sterling, Theodor D. "Sources of Contextual Constraint Upon Words in Sentences," *J. Exp. Psychol.*, vol. 57, no. 3, 1959, pp. 171-179.

23. Edwards, Allen L. *Experimental Design in Psychological Research*, Holt, Rinehart & Winston, Inc., New York, 1960.

24. French, Norman R., Carter, Charles W., Jr., and Koenig, Walter, Jr. "The Words and Sounds of Telephone Conversation," *The Bell System Technical J.*, vol. 9, 1930, pp. 290-324.

25. Miller, George A., and Chomsky, Noam. "Finitary Models of Language Users," in *Handbook of Mathematical Psychology*, vol. 1, R.D. Luce, R.R. Bush, and E. Galanter (eds.) John Wiley & Sons, Inc., New York, 1963, pp. 419-491.

26. Harris, Zellig S. "Co-occurrence and Transformation in Linguistic Structure," *Language*, vol. 33, 1957, pp. 283-340.

27. Harris, Zellig S. "Linguistic Transformations for Information Retrieval," in *Proceedings of the International Conference on Scientific Information (1958)*, National Academy of Sciences-National Research Council, Washington, 1959, pp. 937-950.

28. Siegel, Sidney. *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Co. Inc., New York, 1956.

29. Lewis, Don. *Quantitative Methods in Psychology*, McGraw-Hill Book Co. Inc., New York, 1960.

30. Johnson, G.R. "An Objective Method of Determining Reading Difficulty," *J. Educ. Res.*, vol. 21, 1930, pp. 283-287.

31. Zipf, George K. *The Psycho-Biology of Language*, Houghton Mifflin, Boston, 1935.

32. Hall, John F. "Learning as a Function of Word Frequency," *Amer. J. Psychol.*, vol. 67, 1954, pp. 138-140.

33. Sumby, William H. "Word Frequency and Serial Position Effects," *J. Verb. Learn. Verb. Behav.*, vol. 1, no. 6, May 1963, pp. 443-450.

34. Lloyd, Kenneth E. "Short-Term Retention as a Function of Word Frequency," *J. Verb. Learn. Verb. Behav.*, vol. 3, no. 4, August 1964, pp. 286-289.

35. McGinnies, Elliott M., Comer, Patrick B., and Lacey, Oliver L. "Visual-Recognition Thresholds as a Function of Word Length and Word Frequency," *J. Exp. Psychol.*, vol. 44, no. 2, 1952, pp. 65-69.

36. Newbigging, P. Lynn, and Hay, J.M. "The Practice Effect in Recognition Threshold Determinations as a Function of Word Frequency and Length," *Canad. J. Psychol.* vol. 16, 1962, pp. 177-184.

37. Lyon, D.O. *Memory and the Learning Process*, Warwick & York, Inc., Baltimore, 1917.

38. Thurstone, Louis L. "The Relation Between Learning Time and Length of Task," *Psychol. Rev.*, vol. 37, 1930, pp. 44-53.

39. Hovland, Carl I., and Kurtz, Kenneth H. "Experimental Studies in Rote-Learning Theory: X. Pre-Learning Syllable Familiarization and the Length-Difficulty Relationship," *J. Exp. Psychol.*, vol. 44, no. 1, 1952, pp. 31-39.

40. Bentley, M. "Tools and Terms in Recent Researches: A Noun Modifies a Noun Modifies a Noun Modifies a Noun," *Amer. J. Psychol.*, vol. 57, 1944, pp. 268-269.

41. Sanford, Fillmore H. "CP and Rules for English Usage," *Contemp. Psychol.*, vol. 8, no. 3, 1963, pp. 92-93.

42. Francis, W. Nelson. *The Structure of American English*, Ronald Press Company, New York, 1958.

43. Nida, Eugene A. *A Synopsis of English Syntax*, Summer Institute of Linguistics, University of Oklahoma, Norman, 1960.

44. Gammon, E.R. "On Representing Syntactic Structure," *Language*, vol. 39, 1963, pp. 369-397.

45. Yngve, Victor H. "The Depth Hypothesis," in *Structure of Language and Its Mathematical Aspects*, R. Jacobson (ed.), American Mathematical Society, Providence, R.I., 1961, pp. 130-138.

46. Lees, Robert B. "Comments: On Hockett's Paper," in *Structure of Language and Its Mathematical Aspects*, R. Jacobson (See Reference 45.)

47. Lambert, Wallace E., and Paivio, Allan U. "The Influence of Noun-Adjective Order on Learning," *Canad. J. Psychol.*, vol. 10, 1956, pp. 9-12.

48. Paivio, Allan U. "Learning of Adjective-Noun Paired Associates as a Function of Adjective-Noun Word Order and Noun Abstractness," *Canad. J. Psychol.*, vol. 17, 1963, pp. 370-379.

49. Woodworth, Robert S. *Experimental Psychology*, Henry Holt and Company, Inc., New York, 1938, p. 716.

50. Mehler, Jacques. "Some Effects of Grammatical Transformations on the Recall of English Sentences," *J. Verb. Learn. Verb. Behav.*, vol. 2, no. 4, November 1963, pp. 346-351.

51. Gough, Philip B. "Grammatical Transformations and Speed of Understanding," *J. Verb. Learn. Verb. Behav.*, vol. 4, no. 2, April 1965, pp. 107-111.

52. Bousfield, W.A., and Cohen, B.H. "The Occurrence of Clustering in the Recall of Randomly Arranged Words of Different Frequencies-of-Usage," *J. Gen. Psychol.*, vol. 52, First Half, January 1955, pp. 83-95.

53. Gorman, Aloysia M. "Recognition Memory for Nouns as a Function of Abstractness and Frequency," *J. Exp. Psychol.*, vol. 61, no. 1, January 1961, pp. 23-29.

54. Underwood, Benton J., and Postman Leo. "Extraexperimental Sources of Interference in Forgetting," *Psychol. Rev.*, vol. 67, no. 2, March 1960, pp. 73-95.

55. Waugh, Nancy C. "Free Versus Serial Recall," *J. Exp. Psychol.*, vol. 62, no. 5, 1961, pp. 496-502.

56. Winnick, Wilma A., and Kressel, Kenneth. "Tachistoscopic Recognition Thresholds, Paired-Associate Learning, and Free Recall as a Function of Abstractness-Concreteness and Word Frequency," *J. Exp. Psychol.*, vol. 70, no. 2, 1965, pp. 163-168.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Human Resources Research Office, The George Washington University, Alexandria, Virginia 22314	2a. REPORT SECURITY CLASSIFICATION Unclassified; 2b. GROUP
3. REPORT TITLE EFFECTS OF GRAMMATICAL FACTORS AND AMOUNT OF MATERIAL ON MEMORIZING PARAGRAPHS, SENTENCES, AND WORD LISTS	
4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)* Technical Report	
5. AUTHOR(S) *(First name, middle initial, last name)* Joseph F. Follettie, and Ann F. Wesemann	
6. REPORT DATE June 1967	7a. TOTAL NO. OF PAGES 68; 7b. NO. OF REFS 56
8a. CONTRACT OR GRANT NO. DA 44-188-ARO-2; b. PROJECT NO. 2J014501B74B 02; c.; d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report 67-9; 9b. OTHER REPORT NO.(S) *(Any other numbers that may be assigned this report)*
10. DISTRIBUTION STATEMENT Distribution of this document is unlimited.	
11. SUPPLEMENTARY NOTES Study of Factors That Make Written Material Difficult to Learn	12. SPONSORING MILITARY ACTIVITY Office, Chief of Research and Development, Department of the Army, Washington, D.C. 20310

13. ABSTRACT

The effects of certain measurable characteristics of written material upon speed of memorization were studied. These characteristics were of four general classes: (a) load factors reflecting informational density of a selection, (b) length of the selection, in grammatical units of various kinds, (c) factors dealing with alternative, grammatically equivalent ways for presenting the same semantic material, and (d) word frequency factors. Of the four types of stimulus characteristics studied, length factors were found to have the clearest and most potent effects upon rote memorization. (Author)

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Memorization						
Recall						
Meaningful Material						
Grammatical Factors						
Comprehension						
Readability						
Psycholinguistics						

Unclassified
Security Classification

DISTRIBUTION LIST

1 DIR HEL APG MD
2 ENGNR PSYCHOL LAB PIONEERING RES DIV ARMY NATICK LABS NATICK MASS
4 TECH LIB ARMY NATICK LABS NATICK MASS
3 CO ARMY CHEM R+D LABS EDGEWOOD ARSNL MD ATTN LIBN
1 CG ARMY ELEC PG FT HUACHUCA ATTN TECH LIB
1 CHF DEPT OF CLIN & SOC PSYCH WALTER REED ARMY INST OF RES WASH, D.C.
1 CO DUGWAY PG UTAH ATTN TECH LIB
1 DEF SUPPLY AGY CAMERON STATION ATTN LIB
3 ARMY WAR COLL CARLISLE BKS ATTN LIB
1 US MILIT ACAD WEST POINT ATTN LIB
1 COMDT ARMY AVN SCH FT RUCKER ATTN SCH LIB
2 COMDT ARMY SECUR AGY TNG CTR + SCH FT DEVENS ATTN LIB
2 COMDT NATL WAR COLL FT LESLEY J MCNAIR ATTN CLASSF RECORDS BR LIB
1 MED FLD SERV SCH BROOKE ARMY MED CTR FT SAM HOUSTON ATTN STIMSON LIB
1 ARMY QM SCH FT LEE ATTN LIB
1 ASST COMDT ARMY AIR DEF SCH FT BLISS ATTN CLASSF TECH LIB
2 HQ ABERDEEN PG ATTN TECH LIB
1 CLIN PSYCHOL CONSLT OFC OF CHF PSYCHIAT + NEUROL CONSLT OFC OF SURG GEN ATTN LT COL MSC
50 COMDR ODC CAMERON STA VA
1 U S ARMY BEHAVIORAL SCI RES LAB WASH, D.C. ATTN CRD-AIC
1 ARMY ABN ELEC + SPEC WARFARE BD FORT BRAGG ATTN LIB
2 CO HQ ARMY BROADCASTING + VISUAL ACTVY PACIFIC APO 331 SAN FRAN ATTN PUBS PROC OFC
2 CHF AUDIO VISUAL APPLICAT OFC ARMY PICTORIAL DIV OFC OF CHF SIG OFCR
1 DIR ARMY LIB PENTAGON
1 PERS + TNG BR PSYCHOL SCI DIV OFC OF NAV RES
2 HUMAN FACTORS DEPT COMM PSYCHOL DIV NAV TNG DEVICE CTR PT WASHINGTON
1 SCI INFO EXCH WASHINGTON
2 RESEARCH ANALYSIS CORP MCLEAN VA 22101
2 U OF PGH LEARNING R+D CTR ATTN DIR
1 HUMAN SCI RES INC MCLEAN VA
3 CTR FOR RES IN SOCIAL SYS AMER U ATTN LIBN
2 EDUC + TNG CONSULTANTS LOS ANGELES ATTN PRINCIPAL SCI
1 DIR CTR FOR RES ON LEARNING + TEACHING U OF MICH
1 EDITOR TNG RES ABSTR AMER SOC OF TNG DIRS U OF TENN
1 CTR FOR RES IN SOCIAL SYS AMER U
2 AMER INST FOR RES SILVER SPRING
1 TUFTS U HUMAN ENGNR INFO + ANLS PROJ
1 HUMAN FACTORS RES GP WASH U ST LOUIS
1 AMER PSYCHOL ASSOC WASHINGTON ATTN PSYCHOL ABSTR
1 AMER BEHAV SCI CALIF
1 NY STATE EDUC DEPT ABSTRACT EDITOR AVCR